Web and Philosophy, Why and What For?

Alexandre Monnin, Harry Halpin, Leslie Carr

PhiloWeb 2012

Web and Philosophy: Why and What For?

Proceedings of the WWW2012 conference workshop PhiloWeb 2012

"Web and Philosophy, Why and What For?"

Lyon, France, Apr 17, 2012.

Edited by:

- Alexandre Monnin
  (Université Paris 1 Panthéon-Sorbonne, Equipe Phico, composante EXeCO, Paris, France ; Institut de Recherche et d'Innovation du Centre Pompidou, Paris, France ; INRIA, Equipe Wimmics, Sophia Antipolis, France).

- Harry Halpin
  (Institut de Recherche et d'Innovation du Centre Pompidou, Paris, France).

- Leslie Carr
  (University of Southampton, Southampton, UK).

CEUR Workshop proceedings - Vol-859 - urn:nbn:de:0074-859-9

Session 1: Web and Society

1. A Methodology for Internal Web Ethics (Slides)
   Michalis Vafopoulos, Petros Stefaneas, Ioannis Anagnostopoulos, Kieron O'Hara
2. What Social Ontology for Social Web? An Assemblage Theory Promoted
   Eddie Soulier

Session 2: Web and Cognition

3. Filter Bubble and Enframing: On the Self-Affirming Dynamics of Technologies (Slides)
   Andreas Beinsteiner

Session 3: Web and Language

4. Where Do 'Ontologies' Come From? Seeking for the Missing Link (Slides)
   Aurélien Bénel
5. Web Metaphysics Between Logic and Ontology (Slides)
   Philippe Lacour

Session 4: Web and Logic (& Computation)

6. CoReWeb: From Linked Documentary Resources to Linked Computational Resources (Slides)
   Alexandre Monnin, Nicolas Delaforge, Fabien Gandon
7. Containing the Semantic Explosion (Slides)
   Jaimie Murdock, Cameron Buckner, Colin Allen

A methodology for internal Web ethics

Michalis Vafopoulos
National Technical University of Athens
9 Heroon Polytechneiou st.
15773, Zografou Campus, Greece
+302107722538
vaf@aegean.gr

Petros Stefaneas
National Technical University of Athens
9 Heroon Polytechneiou st.
15773, Zografou Campus, Greece
+302107721869
petros@math.ntua.gr

Ioannis Anagnostopoulos
University of Central Greece
2-4 Papasiopoulou st.
35100, Galaneika – Lamia, Greece
+302231066937
janag@ucg.gr

Kieron O'Hara
University of Southampton
University campus, Southampton
SO17 1BJ, United Kingdom
+4402380592582
kmo@ecs.soton.ac.uk

ABSTRACT
The vigorous impact of the Web in time and space arises from the fact that it motivates massive creation, editing and distribution of information by Users with little knowledge. This unprecedented continuum provides novel opportunities for innovation but also puts under jeopardy its survival as a stable construct that nurtures a complex system of connections. We examine the Web as an ethics determined space by demonstrating Hayek's theory of freedom in a three-leveled Web: technological, contextualized and economic. Our approach accounts for the co-dependence of code and values, and assumes that the Web is a self-contained system that exists in and by itself. This view of internal Web ethics directly connects the concept of freedom with issues like centralization of traffic and data control, rights on visiting log file, custom User profiles and the interplay among function, structure and morality of the Web. It is also demonstrated, in the case of Net Neutrality, that generic freedom-coercion trade-offs are incomplete in treating specific cases at work.

Categories and Subject Descriptors
WSSC: "webscience.org/2010/E.4.3 Ethics in the Web"

Keywords
Web ethics, freedom, economic Web, contextualized Web, centralization of traffic and data control.

1. INTRODUCTION
The amount of information on the Web is growing exponentially. Only in YouTube, 48 hours of video are uploaded every minute or nearly 8 years of content per day. Users' demands for a fast, secure, reliable, all-inclusive, trustworthy and general-purpose Web are uncontrollable. In 2010, the top 10 Web sites accounted for about 75 percent of all US traffic, compared to the 31% in 2001. Business controversies on issues like the monetization of links and excessive market power in searching and mobile applications are coming to the fore, whilst novel business models are changing the market rules (e.g. peer production, crowdfunding). Some executives and interest groups are still trying to conquer the Web by limiting the freedom to connect and update its content. Controversies have been also transferred to the legal battlefields. Contentious legal initiatives (e.g. SOPA) are causing both small and gigantic power games among governments, industries and non-governmental organizations. Concerns about identity, privacy and security are more often in the headlines. Although technically right solutions exist, these are have not been adopted yet (e.g. PKI, P3P, eID). HTML5 seems to gain interest well beyond technological outsets, and Open Data initiatives are revolutionizing science, business and government. Diverse debates and discussions are indirectly or directly connected to the Web ecosystem and outspread across the social discourse. Symbolically, all these issues are gathered under the rhetoric of online access as an emerging universal human right. Lately, national constitutions have started to incorporate it as a basic right (e.g. Norway). Internet and Web pioneers share different views on the issue, thus driving a creative dialogue about our live with the Web. This dialogue has raised, in various different ways and on as many different occasions, the following question: what kind of Web is more beneficial for society? Surely, as the transformational impact of the Web across society grows, the pressure to define its technological principles and the underlying moral values will escalate. Otherwise, we run the risk of ending up with a restrained, fragmented and autistic Web.

PhiloWeb2012 workshop in 21st World Wide Web Conference'12 (WWW 2012), April 17, 2012, Lyon, France.

2. THE NEED FOR WEB ETHICS
Generic questions about Web's transformational potential have been brought into the agenda of many disciplines. Philosophical thinking and engineering should be in the front line by forming the main questions and setting the research framework. On this campaign single-sided analysis (i.e. technological or social) is not sufficient to tackle these complex and multifarious issues. Domain-specific analysis should be orchestrated by more generic approaches, expanding the solution range. Having defined existence, time and space in the Web [36], the next relevant quest is to consider its moral aspects.

Ethics is the branch of philosophy that deals with the study of good and evil. Its fundamental questions are often repeated through time, adaptive to the historical and social conditions. These fundamental questions include the definition of good and evil, the relation between morality and truth, the limits of freedom of will, the definition of right and wrong etc. Applied ethics is the branch of philosophy concerning the application of ethics to specific problems or classes of problems. From 1960's till today the field of applied ethics has seen remarkable growth. Business

ethics, biomedical ethics, computer ethics, animal rights and environmental ethics are some of the most active areas in modern applied ethics [21]. The Vietnam War, the great progress in technology, the wide spread of drugs and contraceptives, the degradation of the environment, have raised a series of questions that could not be answered by traditional theories of ethics. An important contribution of applied ethics to the field of law is "A Theory of Justice" [33]. Computer ethics is the branch of applied ethics, which examines the social and ethical impact of information technology [18, 23]. More particularly, it focuses on the social impact that information technology has on our lives, the nature of such impact, and the utilization of technology in an ethical manner. Examples include issues related to the cybercrime, the protection of privacy, copyright and patents, the digital divide, and the use of computers in the workplace. The variety of technological applications creates new and unexpected situations and possibilities, thus causing new ethical dilemmas and values to emerge. For example, protection of personal data by electronic devices is of particular relevance to our society - to remember that only fifteen years ago the relative sensitivity was rudimentary. Lately, information ethics [12, 13] shed new light on many traditional ethical issues in computing.

The last twenty years there is a growing literature on the study of the ethics of the cyberspace encompassing all kinds of interactions among Users and the Internet [14, 31, 35]. Indicative topics include the ethics of blogging, free Speech and anonymity, pornography, censorship, intellectual property, privacy and regulation, spam and advertising, Internet as a medium of communication, accountability and trust, hacking, and the Internet access as a basic human right.

The Web has been built on the Internet stack, enabling the inter-linkage of digital beings. Despite the fact that it shares some common characteristics with its underlying technologies, creates a new feasibility and actuality space. The Web is sufficiently unusual, transformative and necessary to human existence, and as such it requires more systematic philosophical thinking to describe its ethically-relevant properties [28]. Initial motivation behind the development of the Web was based on ethical principles like esteem, pride, excellence, absence of guilt, rewards, and indignation [28]. Originally it was more a closed "Aristotelian world" than a space governed by rules, roles, hierarchies and deliverables. We believe that the above-mentioned virtues are the core driving forces of its exponential impact. These classic values that inspired the inventor and early Web Users and supported its massive dissemination, have now become more specific in practice. For instance, the discussion about freedom of expression incorporates the issue of Net Neutrality and self-determination that is connected to the privacy of online data.

One of the first questions for Web ethics should be a more comprehensive identification of the values that motivated its creation. An open conjecture in this line of inquiry has to do with the question whether different magmas of values and code could initiate similar decentralized information systems. Another question is how these evolving values affect the impact of Web in diverse social contexts and under what sort of prerequisites they can be sustainable.

It is now the time for scholars to look deep in the heart of the Web creation, to propose and engineer perspicacious solutions that will benefit the entire society. The quest for new requirements should directly address the needs, and promote human values. Web ethics should be thoroughly investigated in order to become a handy compass for Users, entrepreneurs and governments to direct their decisions towards prosperous ways.

3. INTERNAL WEB ETHICS

Web has been evolved from a piece of software code to a dynamical ecosystem of Users and multi-purpose functionalities. Despite its profound importance, Web ethics is still an unexplored research field. As such, it requires systematic research by determined experts.

The core of our methodology consists of two parts, firstly, the historical evolution of the Web and, secondly, a three-leveled approach thereof as this will be introduced below. The Web in its early stages was meant to address mainly technological needs, such as an interlinked bulleting board with low levels of interaction. In subsequent years, though, the Web evolved and became a construct of multiple interlocking contexts, and was even used to enable financial transactions. Users not only post and link digital content, but also communicate, comment, work, advertise, exchange information and physical goods in and through the Web. The social aspects of the Web are fashioned as the ability to create contexts, and an important part of them, economic contexts. Intense social and economic online transactions result into a dynamic magma of values and code. This fundamental standard implies that Web ethics should be studied under the assumption of inherent codependence between User and System (or equivalently Actor and Network [22]). Also, a sound definition regarding existence, time and space is necessary to describe the moral values tied to the Web as a system [36] In order to focus on our the methodology we propose in this paper, we assume that the Web is the only system existing in the universe ("manna from heaven" hypothesis"). Let us call this methodology the *internal Web ethics* analysis. Our approach extents the Web science perspective, which investigates the Web as a self-standing techno-social artifact [5, 38].

3.1 Magma of Users and code

Till the mass dissemination of Web 2.0, the main point in the ethics of computation took for granted that there was clear distinction between the technological and the social methodologies analyzing related phenomena. Technology was considered an autonomous force that changed society, and its methodology had a simple cause/effect form (technological determinism). Others believe the opposite, i.e., society is an autonomous force that changes technology (social determinism). Web 2.0 created a de facto indissoluble magma of Code and Users (techno-social evolution). Hence, the classic technology-society division is irrelevant in capturing the essence of the active User participation. The evolving interdependence between Code and Users can be addressed by models, which are built on the co-dependence of human moral values and engineering principles.

3.2 Being, time and space in the Web

Applied ethics methodologies refer to well-defined application domains. We believe that for the purposes of Web ethics a suitable framework is the definition of Web space [36]. A theory about existence in time and space is necessary to frame a tractable approach for the moral analysis of the Web. In [36] it has been proposed a notion of existence in the Web, based on a pragmatic definition of Being in general: *"a Being exists if and only if there is a communication channel linking to it"*. Being in the Web implies that the communication channel is concrete, identifiable and visible. Uniform Resource Identifier (URI) is the most profound and stable technology about creating communication channels in the Web. It requires the minimal description of

invariant elements in communication through the Web and acts like the "fingerprint" of the Web Being because it is directly connected to existence (birth, access, navigate, edit and death). Thus, a Web Being is defined as follows: "*Web beings are defined to be Beings that can be communicated through the Web.".* The source of value for Web Beings is concentrated on how *digitality* is mutated by the *linking* potential, enabling them to be anywhere, at anytime. Users are "potential" owners of every Web Being, in the sense that the Being may not reside in the hardware but can be downloaded almost instantly. This expansion of the concept of existence is captured by the concept of *virtualization*, which describes the augmented potentialities of Web Being as a digital and distributable unity. The Web Space could be considered as a division of position and place of online content, created by the links among the Web Beings. Each Web Being is occupying a specific locus in the Web network. Identification in the Web Space is given by the URI namespace. Location is specified by a triplet of URIs, namely the URIs of the Web Being and the incoming and outgoing links. These links provide orientation by acting as a three-dimensional "geographic coordinate system" in the Web. The act of creation or deletion of a Web Being or a link, alters the Web Space. Hence, the evolving Web Space is fully describable by the lifetime processes of Web Beings and links. Except for the "book-keeping" clock time defined in Physics, time could be a series of choices in space. Web time is a series of choices (visits) in the Web Space that can be defined as Bergsonian durations, since visiting selections attach meaning and define casual relationships among Web Beings. This approach of time as duration is characterized by indeterminism, heterogeneity and irreversibility. In the Web, durations are becoming discoverable, observable, traceable, able to process and massive.

3.3 The "manna from heaven"

The study of codependence among Code and Users is really complicated. Initially, we suggest that on the first level the Web can be studied as the only existing system in the world. Human beings are communicating and working solely through and with the Web. A compassionate 'God' provides the necessary quantity of 'manna', fulfilling all human needs, with no cost and effort. This strong and unrealistic assumption will help us to comprehend bottom to top the moral values and their inter-connections to the complex actualities of the engineering principles. The analytical outcome of the first level will prepare us to study the effects of the Web in the entire human society. A characteristic domain of application of this assumption can be Net Neutrality issues. It will include the comparative analysis between established and emerging of new theories in the social, technological and economic domain. Analyzing the internal Web ethics at the first level will provide us with the necessary insights about neutrality as the interplay of core values and the engineering of the Web.

3.4 Technology, context and economy

The Web can be analyzed on three levels: the technological, the contextual and the economic, since they reflect its historical evolution from plain software to living ecosystem. The Web technology is built on the Internet, resulting huge amounts of data created by billions of Users (technology level). On top of this software, various new contexts have expanded initial functionalities. Context, being a set of tasks or a general framework of attitudes, enables Users to extent the range of information exchange and collaborative action, mainly through trust mechanisms (context level). The establishment of beliefs and attitudes regarding the trustworthiness of Users and associated Web Beings enabled the emergence of business models that are based on exchanges – financial or other – among Users (economic level).

Note here a point made by [30] who argues the importance of the distinction between trustworthiness/trust and reliability/reliance. He locates the distinction in the nature of the interactions between trustor and trustee. Where the interactions are 'static', we merely have a case of reliance (as someone may rely on a bridge that has been well-built, or on a clock that is correct). The emergence of trust out of reliance is an important signal for the move up from the technology level.

For Pettit, trust only comes when the interaction is interactively *dynamic* – i.e. trustworthy agents understand they are trusted, and trust gives them additional motive for behaving in a trustworthy manner. He argues on this basis that trust over the Internet (and *ipso facto* the Web) is impossible without supporting offline relationships and information, and therefore impossible on the 'manna from heaven' assumption discussed above. The reason for Pettit's rejection of trust as a possibility in this context is the fluidity of identity online – how could a trustor come to believe that a virtual contact fulfilled the requirements for interactive dynamism?

Without getting too deeply into this issue, [25] moves the focus for trustworthiness away from the trustee's attitude to the trustor, and toward the claims about her intentions, capacities and motivations the trustee makes. In particular, it is an attractive suggestion that the shift from reliability to trustworthiness happens as these claims become less deterministic, more implicit and less precise. There is no exact borderline or tipping point, but this conveys the importance of the agency and the choice for the trustee.

3.5 Hayek's theory of freedom

According to [17], "liberty" or "freedom" is defined to be the absence of coercion of some humans by other humans. This does not mean that one has unlimited options including all physical potentialities of the world. Likewise, it does not account for the internal states of being and any metaphysical notion of freedom or power. The main focus is on the mitigation of coercion as a set of restraints or constraints to human will, imposed by others. As Hayek explains (p.133), *"Coercion occurs when one man's actions are made to serve another man's will, not for his own but for the other's purpose. It is not that the coerced does not choose at all; if that were the case, we should not speak of his "acting."* Similarly, Hayek defines important facets of coercion like deception and fraud, as forms of controlling the information upon which a human counts; this information makes a human do what the deceiver wants him to do. Despite the fact that coercion suggests both the threat of inflicting harms and the intention thereby to cause certain outcomes, it does not necessarily involves all influences that humans can exercise upon the acting of others and acquire full control of the environment. Coercion is undesirable because it *"prevents a person from using his mental powers to the full and consequently from making the greatest contribution that he is capable of to the community."* (p.134). On the contrary, freedom is desirable *"because every individual knows so little and, in particular, because we rarely know which of us knows best that we trust the independent and competitive efforts of many to induce the emergence of what we shall want when we see it."* (p.29). It is freedom that releases the unforeseeable and unpredictable; these little accidents in human behavior, which are so vital for innovation. As Hayek argues (p.31) *"It is because we do not know how individuals will use their freedom that it is so important."* and "*Freedom granted only*

when it is known beforehand that its effects will be beneficial is not freedom." These accidents are the resultant of knowledge and attitudes, skills and habits, formed by human interaction and, in most cases, they do not simply occur but evolve. In order to flourish they must be supported by the existence of complementary concepts like some personal sphere, property, state, rules, competition and responsibility. The emergence of personal sphere and property assists individuals to avoid coercion from others. The only means to prevent coercion is the potential threat tied to coercion. States typically monopolize coercive power. In free societies, the State exercises minimal enforcement of coercive power, which nurtures individual creativity and competitive markets based on just distribution of property and responsible individual behavior. Particularly, *"Since coercion is the control of the essential data of an individual's action by another, it can be prevented only by enabling the individual to secure for himself some private sphere where he is protected against such interference. … It is here that coercion of one individual by another can be prevented only by the threat of coercion assured free sphere."* (p.139). The acquisition of property is the first step towards the limitation of personal sphere and against coercive action. The next steps include the initiation of general rules governing the conditions under which behaviors and attitudes become part of such individual spheres (it is clear that carefully-crafted data protection rules are vital for both steps, which makes the lack of cooperation, or even of an agreed framework, between the EU and the US, not to mention India and China, all the more disturbing). It is crucial to ensure that the range and content of these rules is not determined by the deliberate assignment of particular things to particular persons. *"The decisive condition for mutually advantageous collaboration between people, based on voluntary consent rather than coercion, is that there be many people who can serve one's needs, so that nobody has to be dependent on specific persons for the essential conditions of life or the possibility of development in some direction. It is competition made possible by the dispersion of property that deprives the individual owners of particular things of all coercive powers."* (p.141). The degree of freedom in a society is directly related to the minimal enforcement of coercive power by the state according to general and no discriminative rules and the safeguarding of competitive market conditions. Competition as the existence of an efficient number of alternative offers is fundamental in the case of providing life-critical services. Generally, *"whenever there is a danger of a monopolist's acquiring coercive power, the most expedient and effective method of preventing this is probably to require him to treat all customers alike, i.e., to insist that his prices be the same for all and to prohibit all discrimination on his part. This is the same principle by which we have learned to curb the coercive power of the state."* (p.136). Having argued about the strategic role of state in minimizing coercion does not connotes that individuals enjoy only the opportunity and the burden of choice; it also highlights that individuals must accept the consequences of their choices and the resulting approbation or censure for them. In a free society freedom and responsibility should be interlocked.

4. THE WEB AS A SPACE OF FREEDOM

For many philosophers, freedom is not just one of the values but constitutes the source and prescribes the conditions of most moral values [3]. Hence, a theory about freedom is necessary in order to explore the internal Web ethics. In the present article, Hayek's analysis about freedom is adapted because is adequately consonant to the main architectural principles of the Web artifact, namely: lack of central authority, openness, variety of choices, distributed empowerment of individuals and liberal underpinning. Hayek's approach is not the only theory of freedom that can be used to analyze Web ethics. Its clarity and generality enable us to build a starting point that will be extended and refined with other theories to capture the ethical aspects of the Web.

Freedom creates more options to solve problems collectively and to innovate, but some of these options may be used in ways that cause coercion *("freedom-coercion" tradeoff*). Thus, the question enveloping each theory is how to construct a system that selects, with minimum social cost which positive options to sacrifice in order to minimize coercion (or the dual problem). Hayek's approach could be considered to offer one of the systematic answers in this question. In particular, his theory is briefly transcribed as follows:

- *State* posses the monopoly to enforce coercive power through *General Rules*.
- *Personal Sphere* and *Property* counterweight *state* power.
- *General Rules* are enforced equally and describe the borderlines between *state and Personal Sphere.*
- *Property* is a basic realization of *General Rules.*
- *Competition* is possible by the *dispersion of Property.*
- *Mutually advantageous collaboration* is based on *Competition* in service provision.
- An effective *anti-monopolistic policy* is to require from the monopolist (including the state) to *treat all customers alike*.
- *Individuals* should be *responsible* and *accountable* for their actions.

In the following Subsections we consider some "freedom-coercion" tradeoffs on three levels of abstraction (technology, context, economy) according to Hayek's conceptualizations, in order to gradually build a set of important issues about living with the Web.

4.1 The Technological Web

The Web is an engineered artifact, not some natural phenomenon. It has been created as an Internet application and its building blocks are crafted in software code. In this sense, technological underpinnings are vital for its existence.

4.1.1 Internet infrastructure

Internet has been evolved from communication architecture for computers to generative system for innovative software, basically because it was built on simple principles that transfer the power of choice to equally trusted single Users. The absence of central gate keeping and the unprecedented decentralized power in action is coming with two major costs: (a) inefficient personal identity management and thus, lack of security and (b) not guaranteed quality of transmission.

The notion of Internet freedom is related to the free access and inter-connection of any compatible software code developed by Users over the Internet network. Coercive powers are mainly arising due to badware applications (e.g. computer-zombies), traffic censorship (e.g. "Snooping" - accessing information within Internet packets [4]) and inadequate quality of transmission. Personal sphere for Internet Users is described by their IP address whenever they are connected to Internet. IP addresses are traffic data that can only be processed for certain reasons (e.g. payments). Ordinarily, they are considered by Data Protection Authorities and courts to be personal data, despite the fact that courts in some countries (e.g. France) have reached conflicting decisions [20].

O'Hara has argued that the revolutionary aspect of the World Wide Web is that it is a decentralised information structure. This democratic decentralisation is a key factor in the added value that the Web provides, because it facilitates the serendipitous reuse of information in new and unanticipated contexts. However its basic principle, of free flow of information packets and a very simple set of rules and standards underpinning these complex structures, is being undermined by attempts to restrict information flow. As use of the Web has spread, illiberal regimes feel threatened, but thanks to the hands-off approach of the 1990s, there are no affirmative globally-recognised principles governing the flow of information online. Currently, China is still focusing on a censorship-based approach to information control, using methods in direct opposition to the Web's essential governing principle of decentralisation. The liberalism of the Web has two distinct levels: first, the free flow of information and unrestricted linking helps make the valuable network; secondly the engineering principles of the Web facilitate the efficient flow of information and enables the basic structure to attain balance. In this way, ethical principles (and a strong stand on a political dispute) influence directly even Web infrastructure [26].

4.1.2 The case of Net Neutrality (NN)

The definition of NN and its technical consequences as Internet traffic subject to no hindrances could be further elaborated by using Hayek's ideas. The "first-come first-served" model with no other restriction is extended to Quality of Service (QoS) discrimination as long as there are no special and exclusive contracts at work (limited discrimination and QoS). Hence, in the one hand, no one may have exclusivity to end points, but on the other hand, anyone can pay to have higher QoS in its end point. Alternatively, limited discrimination without QoS tiering can be applied. According to some lawmakers in the US, QoS discrimination is allowed, subject to no particular charge for higher-quality service [10]. The underlying technical challenge is to engineer solutions that ensure NN in combination with higher QoS. This can be achieved by designing Internet infrastructure that allows for implicit traffic differentiation and prioritization of a select traffic, but without any kind of User, network operator or ISP intervention. Such a proposal, which involves an implicit kind of datagram separation rather than an a-priori explicit flow prioritization, is called FAN (Flow-Aware Networking) [19, 34]. FAN may ensure neutrality along with the awareness of QoS [9]. This is because it does not aim to explicitly categorize data flows in distinct classes (e.g. premium, basic), but only to create an occurrence, upon which the implicit separation will be performed solely based on the current link status (e.g. dataflow congestion, traffic bottleneck etc.). Therefore, all datagrams are forwarded unconditionally in the pipeline, but they are also "equal", subject to be separated or even dropped when the network tolerance demands it. The main advantage of FAN-based architectures is that they differentiate the data flow, taking into account only the traffic characteristics of the currently transmitted information. Hence, apart from data discrimination, it is not possible to comprehensively discriminate certain applications, services and end-Users. Such NN-QoS symbiosis does not violate NN and data discrimination principles. It however demands a global implementation approach in infrastructure level, involving common standards in prediction and limitation mechanisms for controlling the quality of transmitted information in the pipeline. The limitation mechanisms may provide a sudden separation of flow, but the decision should be made upon specific network tolerance metrics rather than individual properties of specific flows, such as "who" sends/receives a specific "class" of information.

4.1.3 The Web software

The notion of freedom in the Web software is to freely navigate, create and update Web Beings and links. Its cornerstones are universality, openness and separation of layers in engineering, editing, searching and navigating. [4] argues that "*Keeping the web universal and keeping its standards open help people invent new services."* Coercive powers can be directly injected into the network by Internet infrastructure (e.g. NN). Badware-infected Web Beings [41], central control and censoring of Web traffic are main sources of internal coercion in the Web. The emergence of "walled gardens" in cabled TV and Social Networks [4, 41] are based on isolated or malformed (i.e. without exclusive or open URI) Web Beings that strengthen coercive potential through privacy threats and fragmentation. Furthermore, any effort to manipulate for own benefit the results of indexing and searching processes (e.g. spamdexing [24]) is a form of coercion because it distorts searcheability and navigation.

Navigation in the Web space results in traffic. Web traffic is recorded in the Web Being's log file. Actually, this is the first time that humanity has introduced a universal event log in such a stratified and heterogeneous system. The resulting log file is under *common ownership* by design. Both the Editor who administers and updates the particular Web Being and the Navigator, who visits it, share the same information about this event. Although, the Editor has direct access to the log file residing in the Web server, the Navigator should install particular software to process the source file of his visiting history. Thus, this log file is the core architectural element that manifests the co-operative nature of the Web artifact and should be further analyzed. For the moment, legal and illegal cookies are censoring our moves with or without our consent [1] and "toolbar" applications exchange their services for recording all our navigation history.

During the first Web era, the majority of Users were Navigators and just a small portion of them was editing the Web. At the current Web 2.0 era, 70% of Users are both Navigators and Editors, who can easily edit, interconnect, aggregate and comment upon text, images and video. The underlying structure of the Web graph is characterized by four major characteristics: 1) on-line property (the number of Web Beings and links changes with time), 2) power law degree distribution with exponent higher than two, 3) small world property (the diameter is much smaller than the order of the graph) and 4) many dense bipartite sub-graphs [6]. In order for the Web to be an advantageous multi-purpose space, it should consist of a critical mass of Web Beings and links in an appropriate structure to facilitate navigation. Intuitively, it should be connected, not fragmented, to ease navigation from any Web Being to the entire network. The analysis of the interplay among functions, subsequent structures and moral values is an open question for internal Web ethics.

Treating all Navigators equally is an engineering principle. It is violated (or enriched) by profile customization. Treating all Editors alike is achieved through open technological standards developed by independent bodies (e.g. W3C). Public and private contribution to these institutions is necessary to sustain open and effective standards. Apart from the first class principles of universality, openness and separation, "quality-related" issues could be relevant to Web freedom if navigation and searching is severely degraded. Despite the fact that the explosion of bits in Web 2.0 increased the number of available Web Beings, incommoded the discovery of meaningful answers. This overload

of unstructured content is partially tackled by Search Engines. Semantically structured data (aka Web 3.0) are engineered to anticipate it through machine-processable meaningful reasoning. The quality of content also includes factors like diversity, credibility, accuracy and informativeness of online content and stability of links.

4.2 The Contextualized Web

The Web became a techno-social space for innovation and inter-creativity because it has been transformed from a bulletin board to a context-aware system. It is not only the number of options the Web is providing, but also it is the quality and the usefulness of these options that matters. The Web context emerges as a bridge in the traditional public-private dichotomy. The *privatized (or publicized) space* arises between the private realm of intimacy and individualism and the public realm of citizenship and active participation for the societal good [29]. On the contrary, in the industrial economy, where consumers are mainly exercising the right to use resources, Web Users exercise the full range of property rights, namely: (1) to use, (2) to form, modify and substantiate (3) to benefit from use and (4) to transfer Web Beings.

Context, as a set of tasks or general framework of roles and attitudes, enables Users to extent the range of information exchange and collaborative action, mainly through trust mechanisms. For instance, in Web 2.0, what Users create is not simply content (e.g. reviews) but context. This new contextual framework emerges through the aggregation and collaborative filtering of personal preferences in massive scale [39]. More importantly, it facilitates connected Users to search and navigate the complex Web more effectively, amplifying incentives for quality. Of course, there are many open issues to be solved such as the fashioning of more effective forms of online identities and trusting processes. According to [25], trust is an attitude toward the trustworthiness of an agent. In our Web-only hypothetical world ("manna from heaven" assumption), agents are the Users who control specific Web Beings. Representations, intentions, capacities, motivations and contexts are established and expressed exclusively by Web technologies. Hence, freedom in the contextualized Web is to establish specific contexts in order to form beliefs and attitudes that some Users and their underlying Web beings are trustworthy. Coercive powers can arise from un-trustworthy technologies and governments, social hacking, badware and malicious representations.

However, it is also important to take account of the bad forms that trust can take [2]. The links between coercion and trust are sometimes uncomfortably close. Note, for example, that when [16] describes his theory of encapsulated trust informed by rational-choice ideas in social science, he argues that *"I trust someone if I have reason to believe it will be in that person's interest to be trustworthy in the relevant way at the relevant time … [and if that person] counts my interests as partly his or her own interests just because they are my interests"* (p.19). What strikes the reader is how close this definition of *trust* is to Hayek's definition of *coercion* quoted earlier.

This brings in Baier's notion of antitrust [2], where trust is harmful to the society at large. In this case the focus is on areas where trust shades into coercion, but it is clear that there are other spheres of life where freedom undermines trust, or allows corrosive examples of trust to emerge – cybercrime is an obvious example, where trust among criminals is essential to prevent police infiltration, and where trust among Web users is exploited by criminals. Baier's expressibility test [2] (pp.123-124) asserts that a trust relation is morally acceptable provided that the trustee may express her motives truthfully; this is an important insight, but it must be vulnerable to Pettit's worry that such expression, in the world we are envisaging, could only be mediated by Web technologies.

Nevertheless, communication is central to establishing trust, as Habermas argued [15], and so the rich connectivity of the Web is bound into its function. Antitrust and coercion may well be prices we have to pay for widespread and beneficial trust (repeating Hayek's point that freedom may at all times produce bad outcomes). The point of a Web ethics is to try to ensure not that antitrust happens, but that it is outweighed by beneficial trust to as great a degree as possible consistent with Hayekian notions of freedom.

4.3 The Economic Web

Most needs are better fulfilled through collective effort. In practice, incentives, capabilities, preferences and realizations of effort are heterogeneous and difficult to be synchronized. A powerful metaphor to achieve synchronization is setting efforts and the products of them under a common valuation scheme, a uniform *numeraire.* This numeraire is money, supported by a set of institutions and practices (e.g. the market). It is far beyond the scope of this paper to analyze related economic theory. We limit ourselves to the reassurance that economizing a system is an important factor for its viability, usability and development. The issues posed in preceding layers could be viewed through the economic aspect (e.g. NN as two-sided pricing [11]). The question is how the above-mentioned freedoms can be efficiently engineered and disseminated across Users in particular techno-social contexts.

The Web has not emerged as a business project with hierarchical structures. It has been crafted as a creative and open space of volunteers, predominantly outside traditional market and pricing systems. In our point of view, markets would have never invested such amounts in labor costs to develop this gigantic system. But to be fair, market mechanisms provided the necessary motives and tools to initiate a high-risk idea like Web. Furthermore, the lack of direct compensation and the temporal disconnection between effort and rewards are the shared characteristics among Peer, Procurement and Patronage production models. In the Web, Peer production has been established as a basic form of production, extending David's taxonomy [7] with the fourth P [37] .

The explosion of Web Users occurred as a result of symbiosis between non-financial and financial incentives [37]. Accordingly, freedom in the economic Web pertains to the removal of any possible barrier to economize. Each User should be allowed to apply any business models. Apart from the preceding levels, coercive powers are coming from two economy-related sources: the concentration of power in a minority of Web Beings and Users and the inability of some Users to benefit from the Web economy. As the economic Web grows, state faces unprecedented and complex trade-offs between private interest and social welfare. Three of those are referred as the "Link economy", the "App economy" and the excessive market power in Search Engine market. Recently, the formation of links, a fundamental characteristic of the Web, became the center of business controversies. As traditional content creators (e.g. TV) are losing a large part of their revenue streams from User-Generated substitutes (e.g. micro-blogs), the need for the institution of regulation issues in free reference linking appears. On the other hand, it is argued that Search Engines create exploitable traffic for content creators and that all online content must be open, with

permanent links, so that it may receive in-links, since links are a key to securing efficiency in creating and finding information. However, the economic implications of reference links on attention and revenue have not been analyzed yet, despite their influence over consumer's utility, competition and social welfare. [8] concluded that: *"link equilibria often do not form, even though their formation can lead to higher aggregate profits and better content. This, in the view of the authors constitutes a negative side-effect of the culture of "free" links that currently pervades the web…"*

Despite the fact that Web 2.0 multiplied the pool of Users and content, the direct use of Web technologies has become shallower. Contrastingly to early stages of Web's inception, modern Users are mainly using the Web through established services (e.g. Search Engine, Social Network) and not directly, for instance, by creating their homepage or concentrating and controlling personal data in a privately owned domain. [4] reasons that the tendency for some companies to develop native applications for specific devices (e.g. "app stores") instead of Web applications sterilizes and fragments the Web. [32] demonstrates that the already large levels of concentration in the Web search market are likely to continue. He argues that since the market mechanism cannot provide socially optimal quality levels, there is space for regulatory engagements which may involve the funding of basic R&D in Web search, or more drastic measures like the division of Search Engines into "software" and "service" parts. It seems that massive use is coming with the cost of *centralization* of both traffic and data control. The balance point between innovation coming from large Web companies and innovation from single Users or voluntary groups should be thoroughly examined. In our point of view, this fast evolving centralization is directly against the core values of the Web ecosystem and must be addressed in the direction of transferring back to individual User part of data control. This can be achieved through technologies and business practices that are transparently enabling the User to process and economize personal data. In this campaign, the primer difficulties arising from the fact that now the Web is partly governed by economic forces and traditional institutions, which are characterized by irrelevant or conflicting moral principles. Therefore, one of the fundamental issues for Web ethics is to put this debate to the foreground through the employment of concrete architectural and policy structures (for example, with reference to the conditions, formats and licenses under which Public Sector Information for reuse is made available to citizens).

5. RESULTS AND DISCUSSION

We believe that the Web engineering principles are ethically-relevant and they should be systematically analyzed as such, in order to realize their potential in promoting human values. Web ethics raises the question about what could be a better future with the Web and how we can engineer it. As an emerging field of applied ethics, it discerns the core values of Web's inception and their evolution process in diverse social contexts. Our main arguments are based on the codependence of code and values. The Web is seen as a new form of existence [36] and it is assumed that it is the only existing system. The proposed methodology gradually analyzes the Web's complex reality by enriching underlying technology with human behavior aspects. Our three-levels analysis (technology, context and economy) reflects the historical evolution of the Web from software to a social ecosystem. As the concept of freedom is a prerequisite of most of the moral values, we introduce our methodology on internal Web ethics by demonstrating Hayek's theory of freedom in the three-levels analysis of the Web. We choose Hayek's approach because it reflects nicely the codependence among the architectural engineering principles of the Web and moral values. This correspondence can be summarized as follows:

- centralization of traffic and data control, rights on visiting log file, custom User profiles and interplay among functions, structures and moral values are directly connected to the quality of freedom in the Web,
- issues about freedom in lower levels of the Web ecosystem (i.e. technology) have crucial impact on the subsequent levels of higher complexity (i.e. context, economy) and
- generic freedom-coercion trade-offs are useful in framing the feasibility space but incomplete in treating more specific cases in practice (e.g. NN).

As the Web grows, it becomes essential to balance the need for efficient efforts and the stimulus for more competitors in creating and economizing content and search provision. A basic prerequisite in this effort is to identify and engineer its core moral values in order to account for an extensive range of User functionalities and pervasiveness in social discourse. This ongoing work can be further inspired by philosophical theories and historic periods [27] (pp.207-209). Also, it will be placed and compared with regards to relevant research about the interplay between technology and society. Providing deeper insights in Web ethics requires the supplementary specification of the suggested model with sound theoretical foundations and more realistic assumptions. Therefore, the next steps should include the enrichment of contextualized Web with theories and technologies about identity, privacy and trust. The study of the ethics of the economic Web should be extended to the study of inequality and distribution theories and detailed business models. During the next phase of this research project, the "manna from heaven" assumption will be relaxed and the three-levels model will be augmented by a fourth level to capture Web's interaction with other real systems. At a latter stage the Web ethics should be able to address more pragmatic questions like: "Can the Web protect itself as a liberal society? How do we manage online identities ethically? How can I deal fairly with people if I don't know their expectations? If I don't even know they are people? " [28]. How the Web's function, structure and evolution are affected by ethics?

The Web is a unique piece of technology not only because of its breakthrough technological innovation, but mainly because it provides a new basis for expressing human creativity, and reveals "inactive" parts of human nature. Apart from understanding its morality, it is an inspiring challenge to transfuse the essence of our experience and the values of the Web to reassess concepts like freedom, choice, participation, inequality and development. We agree with [40] that *"It is not just information that must be free, but the knowledge of how to use it. The test of a free society is not the liberty to consume information, nor to produce it, nor even to implement its potential in private world of one's choosing. The test of a free society is the liberty for the collective transformation of the world through abstractions freely chosen and freely actualised."* The role of Web ethics could be to elaborate and specify the motives and engineering of this new version of utopia.

6. ACKNOWLEGMENTS

We would like to thank Sophia Papaioannou for her useful feedback on an earlier version of this paper.

7. REFERENCES

[1] d' Aquin, M. et al. 2010. Personal monitoring of web information exchange: Towards web lifelogging. *Web Science 2010: Extending the Frontiers of Society On-Line* (2010).

[2] Baier, A. 1994. *Trust and antitrust*. Harvard University Press, Cambridge MA.

[3] Bay, C. 1965. *The structure of freedom*. Stanford University Press, Stanford, California, USA.

[4] Berners-Lee, T. 2010. Long Live the Web: A Call for Continued Open Standards and Neutrality. *Scientific American, November*. 22, (2010).

[5] Berners-Lee, T. et al. 2006. A Framework for Web Science. *Foundations and Trends® in Web Science*. 1, 1 (2006), 1-130.

[6] Bonato, A. 2005. A survey of models of the web graph. *Combinatorial and algorithmic aspects of networking* (2005), 159–172.

[7] David, P. 1992. Knowledge, property, and the system dynamics of technological change. *World Bank Research Observer*. 7, (1992), 215–215.

[8] Dellarocas, C. and Katona, Z. 2010. Media, aggregators and the link economy: Strategic hyperlink formation in content networks. *Working Papers*. (2010).

[9] Domzal, J. et al. 2009. Qos-aware net neutrality. *First International Conference on Evolving Internet, 2009 (INTERNET'09).* (2009), 147–152.

[10] Dorgan, B. 2007. A bill to amend the Communications Act of 1934 to ensure net neutrality. *Library of Congress Bill S.* (2007).

[11] Economides, N. and Tåg, J. 2007. Net neutrality on the Internet: A two-sided market analysis. *SSRN eLibrary*.

[12] Floridi, L. 2010. 5 Information ethics. *The Cambridge Handbook of Information and …*. (2010).

[13] Floridi, L. 2005. Information ethics, its nature and scope. *ACM SIGCAS Computers and Society*. (2005).

[14] Godwin, M. 2003. *Cyber rights: Defending free speech in the digital age*. MIT press.

[15] Habermas, J. 1984. *The Theory of Communicative Action*. Polity Press, Cambridge.

[16] Hardin, R. 2006. *Trust*. Polity Press, Cambridge.

[17] Hayek, F.A. 1960. *The constitution of liberty*. University of Chicago Press Chicago.

[18] Johnson, D. 2003. *Computer ethics*. Prentice Hall.

[19] Kortebi, A. et al. 2004. Cross-protect: implicit service differentiation and admission control. *High Performance Switching and Routing, 2004. HPSR. 2004 Workshop on* (2004), 56-60.

[20] Kuner, C. et al. 2009. *Study on Online Copyright Enforcement and Data Protection in Selected Member States*.

[21] LaFollette, H. 2002. *Ethics in Practice: an anthology*. Wiley-Blackwell.

[22] Latour, B. 2011. Networks, Societies, Spheres: Reflections of an Actor-Network Theorist. *International Journal of Communication*. 5, (2011), 796-810.

[23] Moor, J. 1985. What is Computer Ethics. *Metaphilosophy*. (1985).

[24] Ntoulas, A. et al. 2006. Detecting spam web pages through content analysis. *Proceedings of the 15th international conference on World Wide Web* (2006), 83–92.

[25] O'Hara, K. 2012. *A General Definition of Trust*.

[26] O'Hara, K. 2009. "Let a Hundred flowers blossom, a hundred schools of thought contend": ": Web Engineering in the Chinese Context. *China's Information and Communications Technology Revolution: Social changes and state responses*. Y. Zheng, X. Zhang, ed. Routledge. 121-135.

[27] O'Hara, K. 2010. *The Enlightenment: A Beginner's Guide*. Oneworld, Oxford.

[28] O'Hara, K. 2010. The Web as an Ethical Space. *Proceedings of the British Library Workshop on Ethics and the World Wide Web, 16th November 2010, British Library, London.* (2010).

[29] O'Hara, K. and Shadbolt, N. 2008. *The spy in the coffee machine: the end of privacy as we know it*. Oneworld, Oxford.

[30] Pettit, P. 2008. Trust, Reliance, and the Internet. *Information technology and moral philosophy*. (2008), 161.

[31] Ploug, T. 2009. *Ethics in cyberspace: how cyberspace may influence interpersonal interaction*. Springer-Verlag New York Inc.

[32] Pollock, R. 2010. Is Google the Next Microsoft: Competition, Welfare and Regulation in Online Search. *Review of Network Economics*. 9, 4 (2010), 4.

[33] Rawls, J. 1999. *A theory of justice*. Belknap Press.

[34] Roberts, J.W. and Oueslati, S. 2000. Quality of Service by Flow Aware Networking. *Phil. Trans. Royal Society London*. 358, (2000), 2000.

[35] Spinello, R.A. 2011. *Cyberethics: Morality and law in cyberspace*. Jones & Bartlett Publishers.

[36] Vafopoulos, M. 2012. Being, space and time in the Web. *Metaphilosophy (forthcoming)*. (2012).

[37] Vafopoulos, M. 2012. The Web economy: goods, users, models and policies. *Foundations and Trends® in Web Science (forthcoming)*. (2012).

[38] Vafopoulos, M. 2011. Web Science Subject Categorization (WSSC). *Proceedings of the ACM WebSci'11* (2011), 1-3.

[39] Vafopoulos, M. and Oikonomou, M. 2012. Recommendation systems: bridging technical aspects with marketing implications. *Studies in Computational Intelligence (forthcoming)*. I. Anagnostopoulos et al., eds. Springer.

[40] Wark, M.K. 2004. *A hacker manifesto*. Harvard University Press.

[41] Zittrain, J. 2008. *The future of Internet and how to stop it*. Yale University Press.

What Social Ontology for Social Web? An Assemblage Theory promoted

Eddie Soulier
Université de Technologie de Troyes /
Tech-CICO
12, rue Marie Curie, BP 2060, 10010
TROYES CEDEX, FRANCE
33 (0)6 84 23 55 84
eddie.soulier@utt.fr

ABSTRACT
In this paper, we propose to investigate three main social ontological schools – practice theory, assemblages theories and actor-network theory – to focus on the assemblage theory, and to clarify social ontology underlying the last one, for finally discussing its contribution in the debate on the philosophies of the (social)Web, in order to specify what social ontology for social web.

Keywords
Social ontology, assemblage theory, actor-network theory, practice theory, social web.

1. WHAT IS SOCIAL ONTOLOGY?
Online social interactions form the basis of much online activity including all what is actually important for people: expression of self, social networking, citizen participation, content sharing or, more concretely, all practical aspects of everyday life (online shopping, traveling, health, education, sport, jobs, religion…). As we referred more and more often to the Web as the *social* Web and more explicitly as a set of social relations that link people through the World Wide Web [1] it becomes necessary to go more thoroughly into what "social" is.

If we except the two main theoretical traditions in Social science, as Giddens [2] coins them - holist or objectivist theories like functionalism, systems theory and structuralism on one side, individualist or subjectivist theories like hermeneutic and the many interpretative theories (including rational action theory which must be treated separately, i.e. Coleman [3]) on the other side – the eighties are characterized by some attempt of overcoming the antinomies between 'action' (or agency) and 'structure', which existed in both sociological and philosophical literature. The notions of agency and structure presuppose one another and social theorists proposed different conceptualizations from the relation of the actor to the system which transcend this opposition. The new object of the social sciences becomes something which "stands in the middle", neither the consciousness or experience of individual nor the societal totalities.

The question of what social reality is made of is not only an epistemic issue but also (and more deeply) an ontological one. The emergence of a vivid ontological debate around social ontology is a recent one, stimulate by the work of Margaret Gilbert [4], John Searle [5, 6], Barry Smith [7] or Pierre Livet and Frédéric Nef [8] in France. Nevertheless, we consider that the questioning on the existing entities that the Social sciences speaks about must not be an exclusive privilege of some philosophers or sociologists who position mainly on social ontology, but also the fact of some social scientists or philosophers who wants to renew traditional ontological divisions for better addressing the duality agency-structure of the Social. In doing so, many other social scientists (but not so much) can be invited in the social ontological debate.

Seen like that, and if we willingly exclude for the present discussion the Phenomenologist or Interactionist perspectives[1] due to the limited ambition of this communication, three main frameworks with many variants seem to appear at that time (the eighties): practice theory, actor-network theory and what we propose to call an "assemblage theory", following DeLanda [9]. What we propose to do is to recall these three schools, to focus on the assemblage theory, and to clarify social ontology underlying the last one, for finally discussing its contribution in the debate on the philosophies of the (social)Web.

To be fully honest humanistic, phenomenological or interactionist approaches, as well as the text-centered hermeneutic models tend to adopt what Quentin Meillassoux {10] call in prominent book correlationism[2]. It is why we exclude them.

[1] Like Goffmanian microsociology, ethnomethodology or Conversation Analysis.

[2] By 'correlationism' he means "the idea according to which we only ever have access to the correlation between thinking and being, and never to either term considered apart from the other…[C]orrelationism [indexes] any current of thought which maintains the unsurpassable character of the correlation so defined. Consequently, it becomes possible to say that every philosophy which disavows naïve realism has become a variant of correlationism." (Meillassoux, 2008, p. 5). He denounce the restrained nature (and inconsistency) of the premises of post-Kantian continental philosophy and propose a new, speculative way of philosophising, based on direct access to the external world obtained through mathematical reasoning. Correlationist position tacitly holds that we can't really have experience of the world (or in knowledge of reality) independently of thought or language. The origins of this correlationist turn lie in Kant and his successors, from Husserl to Heiddeger to Derrida.

We claim that it is valuable to pay anew attention to materialist and realist options in social theory and philosophy. So we propose to refer to some contemporary thinkers who perfectly know practice theory as well as actor-network theory or arrangement theories, and who make some effort to extricate social ontology from them. Doing so, we will be in a position to compare the differences between these positive ontologies around few features.

2. SOME MAIN SOCIAL ONTOLOGIES

We retain in this article three social ontologies: practice theory, assemblage theory and actor-network theory.

2.1 Practice Theory

Practice theory has its root in Ancient Greek where the word *praxis* referred to activity engaged in by free men. Aristotle held that there were three basic activities of man: theoria, poiesis and praxis. Three types of knowledge correspond to these three kinds of activities. Aristotle further divided practical knowledge into ethics, economics and politics. New practice theorists which revive Aristotelian conception of moral (and unlike some analytic or libertarians philosophers who try to generate moral consensus on the basis of an ideal of rationality) could be found in philosophical communitarianism, especially in the work of Alasdair McIntyre (*After Virtue*) for who practices is the fabric of virtues or mainly Charles Taylor (*Sources of the Self*). They consider classical liberalism to be ontologically and epistemologically incoherent.

Marx himself also alluded to this concept in his *Theses on Feuerbach* when he stated that "The chief defect of all hitherto existing materialism (…) is that the thing, reality, sensuousness, is conceived only in the form of the object or of contemplation, but not as sensuous human activity, practice (…) All social life is essentially practical (…) Feuerbach, not satisfied with abstract thinking, wants contemplation; but he does not conceive sensuousness as practical, human-sensuous activity". In continuity practice is often also a key entity in critical paradigms as in critical theory ("Frankfurt School" theorists like Jürgen Habermas, e.g his 'Theory of Communicative Action').

The third influential source of contemporary theory of practice is provided by Wittgenstein's *Philosophical Investigations* [11]. Here he rejected the dominant semantic conception of language where "every word has a meaning. This meaning is correlated with the word. It is the object for which the word stands" (Augustine, Confessions, 1. 8.). The exact opposite of what semantic web seems to be! Isn't it? On the contrary, pragmatic conception of language treats language as an activity. His argument is that meaning is use: words are not defined by reference to the objects they designate, nor by the mental representations one might associate with them, but by how they are used.

However, the term "theories of practice" was introduced by Sherry Ortner in the 1980s to refer to recent theorizing in Anthropology, and it is today use to cover the works of a diverse set of philosophers and social theorists [12, 13]. Antony Giddens cited above is himself well-known for his 'theory of structuration' and his holistic view of modern societies. Structuration theory tried to overcome the duality between agency and structure proposing "practice" as a 'Rosetta stone' to comprehend how a social practice both enables and constrains social action. Quite as the same time, Pierre Bourdieu in his *Outline of a Theory of Practice* [14] tried to reconcile the influences of both external social structures and subjective experience on the individual. Michel Foucault [15] is obviously also quoted to belong to the practice camp. In the complementary vein, Michel de Certeau in *The Practice of Everyday Life* [16] develops a theory of the productive and consumptive activity inherent in everyday life. According to de Certeau, everyday life is distinctive from other practices of daily existence because it is repetitive and unconscious.

A 'situated' version of social practice theory was elaborated by Jean Lave [17] who pioneered the theories of situated cognition. She downsized the 'structuralist view' of practice still present in both Bourdieu and Giddens works by emphasizing that knowing is inseparable from doing and by arguing that all knowledge is situated in activity bound to social, cultural and physical contexts. So social practices virtually pre-exist from agent which are 'participants' or, as Andreas Reckwitz [18] coined, 'carriers of the practice' (p. 252). People are always embodied and embedded in ongoing historical processes which belong themself to a socially and culturally structured world. Lave emphasizes the relational interdependence between persons, activity, and world. Social practice theory is thus a theory of relations. The notion of situated activity assumes that subjects, objects, lives, and worlds are made in their relations. "That is, the contexts of people's lives aren't merely containers or backdrops, nor are they simply whatever seems salient to immediate experience. Persons are always embodied, located uniquely in space and in their relations with other persons, things, practices, and institutional arrangements" ([19] p. 2).

Sharing some features with structurationism and some version of social constructivism, she quoted (Hart, 2002, p. 296) saying that: "instead of starting with a presumption of pre-existing bounded entities – whether spatial, social, or individual – a relational approach attends explicitly to ongoing *processes* of constitution. This processual understanding, in turn, is grounded in a theory of praxis that asserts the inseparability of situated practices and their associated meanings and powers relations". So, as Foucault also brings to light, the situatedness of practice involves that living is embedded in political arrangements, hegemonic projects, and diffuse relations of power. But, as other relationisms (in particular Whitehead' one), social practice theory (and more particularly Marxian ones) belongs surprisingly to a relational ontology camp as Hart and Lave reassert, most probably through Hegelian dialectic.

In the new ontological front open by recent disparagements of totality and individuality as basic element of the structure of social life, practice appears as a principal constitutive element – a particular type of entity – in social life[3].

Many studies were done to try to systematize the specific principles and concepts of the various theories of the practice, which still does not form today a real unified theory. *The practice turn in contemporary theory* edited by Theodore R. Schatzki, Karen Knorr Cetina, and Eike von Savigny in 2001 [21] can be viewed as one of the cornerstone of the practice'

[3] Meanwhile at the same time the concept of practice was virulently attack by Stephane Turner [20].

revival[4]. It is said that: "The practice approach … is summed up in its forceful opposition to representational accounts: meaning and language, arising from and tied to continuous activity, cannot be telescoped into representations or mental contents, which themselves acquire the property of being about something by virtue of how people use and react to them" (p. 12). The complexity of practice concept lies in the fact that, as practice theory is absolutely a cultural theory ([18], p. 244), and, as such, *must* to evoke some symbolic stuffs and/or language, although it can't resort to any kind of representation or linguistic features to explain the shared understanding that emerge inside human coexistence, except in term of 'practice'. It argues that even the most apparently "propositional" knowledge acquires meaning only in relation to fields of social and embodied practice.

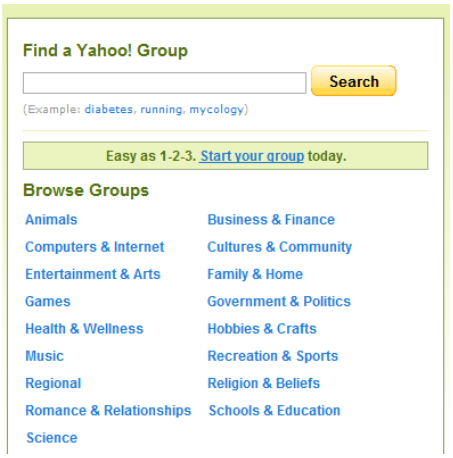

Figure 1. Sameness of practices on Yahoo's site as Aristotle 'praxis' conception. [5]

So, if they are many conceptions of practice, we propose to consider Theodore Shatzki as one of the main philosopher of social sciences who is clearly the most ontological oriented, even if he doesn't cover all practice theorists positions. According to him practice approaches promulgate a distinct social ontology: "the social is a field of embodied, materially interwoven practices centrally organized around shared practical understandings. This conception contrasts with accounts that privilege individuals, (inter)actions, language, signifying systems, the life world, institutions/roles, structures, or systems in defining the social" (Shatzki and al. 2001, p. 3).

2.2 Assemblage Theories

"Assemblage" is a term used by authors of philosophy and sciences to characterize to varying degrees the non-unified, non-hierarchical, non-linear, hybrid, flat, and complex nature of wholes. We refer here especially to the seminal metaphysic work of Whitehead [23], enrich by some postmodern philosophy stances in favour of apparatus [15] and 'agencement' [24], continued by new continental philosophers like Manuel DeLanda [9].

[4] See also *Knowing in Organizations. A Practice-Based Approach* edited by D. Nicolini, S. Gherardi and D. Yanow in 2003 [22].

[5] The only differences relate to 'Computer and Internet'…

Foucault defines an "apparatus" following: "What I'm trying to single out with this term is, first and foremost, a thoroughly heterogeneous set of consisting of discourses, institutions, architectural forms, regulatory decisions, laws, administrative measures, scientific statements, philosophical, moral, and philanthropic propositions – in short, the said as much as the unsaid. Such are the elements of the apparatus. *The apparatus itself is the network that can be established between these elements* (…) By the term "apparatus" I mean a kind of a formation, so to speak, that at a given historical moment has as its major function the response to an urgency. The apparatus therefore has a dominant strategic function (…) I said that the nature of an apparatus is essentially strategic, which means that we are speaking about a certain manipulation of relations of forces, either so as to develop them in a particular direction, or to block them, to stabilize them, and to utilize them. The apparatus is thus always inscribed into a play of power, but it is also always linked to certain limits of knowledge that arise from it and, to an equal degree, condition it. The apparatus is precisely this: a set of strategies of the relations of forces supporting, and supported by, certain types of knowledge" [15] (194-96).

Contrary to Foucault who seeks through the figure of the apparatus to make the idea of structure more dynamic (in structuralism sense), while preserving however the assumption of a certain homogeneity of the elements which are connected, Deleuze will build the differential of the forces which are embodied in assemblages starting from an assumption of radical heterogeneity of their components. "Structures are linked to conditions of homogeneity, but assemblages are not (…) What is an assemblage? It is a multiplicity which is made up of many heterogeneous terms and which establishes liaisons, relations between them (…) Thus, the assemblage's only unity is that of co-functioning: it is a symbiosis, a "sympathy". It is never filiations which are important, but alliances, alloys; these are not successions, lines of descent, but contagions, epidemics, the wind" [25] (p. 69).

The main (but crucial) difference among us is that as Foucault remains in a neo-structualist posture where, in apparatus, entities in relation are linked by internal relations to form a whole as totality (relations of interiority), Deleuze calls 'assemblages' wholes characterized by *relations of exteriority* (DeLanda, 2006, p. 10). Thus, any social entity, on any scale (person, interaction, interpersonal network, City, State…) can be described as an assemblage who offers us a true alternative to organic totalities. In particular, any assemblage could be seen as resulting of an emergence starting from complex interactions between heterogeneous and autonomous components parts.

An assemblage refers therefore to the heterogeneous components which are ordered in any domain of entities, assemblage itself being the system of relations that can be established between these elements. Assemblages are non-essentialist (they are historically contingent actual entities – singular individuals - not instances of ideal forms) and non-totalizing phenomena's (assemblages are not seamless totalities but collections of heterogeneous components that should be analyzed as such). An assemblage is a "multiplicity", a whole made of elements (or parts). Unlike organic totalities an assemblage is a by-product of interactions between components, an emergence. Such multiplicity is a *structure of a possibility space* [26].

Assemblage's identity as possibility space may be "parameterized" (or restricted) along primary axes [27]. A first axis defines the variable roles a component may play: expressive or material. A 'territorializing'/'deterritorializing' axis indicating processes in which a component is involved. These components are defined by relations of exteriority, i.e. their 'role' within a larger assemblage is not what defines them (this would be a relation of interiority). This means that a component is self-subsistent and may be 'unplugged' from one assemblage and 'plugged' into another without losing its identity. Whereas in organic totality the linkages between its components form *logically* necessary relations which make it what whole it is, in an assemblage these relations may be only *contingently* obligatory. This second axis specifies the stability of an assemblage according to the state of its boundaries (sharp and fixed or fuzzy and fluctuating) and the degree of internal homogeneity of its components. The degree of mobility (behavioral factors) of an assemblage may also determine its identity. A third axis defines processes in which specialized information constraints intervene in 'coding'/'decoding' the assemblage. A high degree of territorialization and codification means for an assemblage a weak ability to change. And vice versa an assemblage may be said (relatively or absolutely) decoded and deterritorialized if it is be able to decontextualize a set of relations that partially fixed it (and, thus, destabilized it), rendering them virtual (immanent) and preparing them for more distant actualizations (like communication technology does). Thus the parts of an assemblage are analyzable and assemblage itself has irreducible properties to its parts, without being a 'totality'.

Assemblages are also defined by their tendencies and capacities [27]. Tendencies can make the properties of a whole vary, as when a seed is growing up and changes its own identity, becoming a young plant: here the tendency of any seed is to grow (if nothing prevents it). On the other hand, capacities make a whole exhibits aspects of their identity that were previously hidden, as when an apparently neutral plant turns out to possess unexpected medicinal powers. But tendencies and capacities cannot be listed before they appear due to the relationship between entities component the whole and the all different ways in which they can affected and by affected each other's and by other wholes.

Assemblage theory makes it also possible to position social entities on all scales, from sub-individual to transnational, making the problem of the link between micro- and macro-levels of reality non relevant in this 'flat ontology' perspective. Finally, assemblages necessarily exist in heterogeneous populations, which form their context. The relationship between an assemblage and its components is complex and non-linear: assemblages are formed and affected by heterogeneous populations of lower-level assemblages, but may also act back upon these components, imposing restraints or adaptations in them.

2.3 Actor-Network Theory

Actor–network theory is a recent approach to social theory which originated in the field of science studies. It is known for its controversial insistence on the agency of nonhumans and more generally for claiming the introduction of the too often neglected objects in social sciences. Enlarging the list of workable entities allows envisioning rethinking the old question of blend or alliance among entities with new fresh eyes. Actor-network theory tries to explain how material–semiotic and / or actors-objects networks come together to act as a whole.

Latour pushes thus the concept of assemblage to the limit in the Actor-Network Theory [28). In "*Irreductions*" [29] Latour provides the ontological basis for what has become known as "actor-network theory". Here, an active entity (an agent or actant) is defined neither by itself (identity, essence) nor by its relations (its network). This apparent paradox is possible because the question of the actors and their network is always empirically untied, during trials in which agents, and mediations on which they rely on, operate translations enabling them at the same time (or not) to enter in relation, and to be defined as acting individual and collective entities. The dynamic "mediation-translation-trial" associates (according to dimensions which are themselves heterogeneous) and stabilizes an initial plurality of heterogeneous entities according to a certain trajectory (to go further, see participation).

3. WHAT ARE THE MAIN FEATURES FOR SOCIAL ONTOLOGY?

We try to propose some taxonomy of philosophers in order to see more clearly in some ontological statements, based on a free reading of Harman lecture [30, 31, 32], because he is a sagacious analyst of the current ontological situation.

Harman sets out to develop what he calls an *object-oriented philosophy* (OOP). Taking the tool-analysis as the momentum in twentieth-century philosophy, Harman finds in Heidegger the roots of metaphysics which place the 'things' at the center of reality, like many Heideggerian philosophers (ie Stiegler among many others). Although he considers phenomenology to be deficient in that it subordinates the independent life of objects to our (human) access to them (position closed to Quentin Meillassoux's correlationism). Against the Kantian tradition, his object-oriented philosophy considers the neglected real life of objects to be a 'line of flight' for a new 'speculative metaphysics' [33]. It is possible to pair Whitehead and Latour as object-oriented philosophers according to Harman. The two of them are philosophers of concrete, actual and individual entities (actual entity or actual occasion for Whitehead and 'actant' for Latour[6]). But Harman breaks up at the same time from Whitehead and Latour on the definition of his object ontology. According to him, Whitehead turns entities into clusters of relations, while he hold that *only a non-relational model of object* is capable of accounting for both the transient and enduring faces of reality (Harman, 2011 b., p. 292). Focus on becoming (as well as Whitehead, like Deleuze besides in the current fashion) is only possible with a non-relational ontology (as opposed to what Whitehead or Deleuze would however propose). Compared to the question of becoming only, object oriented philosophy is thus distinguished also from the philosophy of Deleuze, where there is no place for concrete entities[7]. And by rebounds, Deleuze would not be Whiteheadian!

But it is not possible to accord Latour and Harman because Harman [31] promotes a dual category of objects (real objects

[6] It is perhaps insubstantial to qualify Whitehead of philosopher of the process like Rescher [34] does it.

[7] Although bridging between Whitehead and Deleuze was established by Isabelle Stengers.

and sensual objects or intentional objects) completely antagonistic from 'the flat ontology' of Bruno Latour (as Hartman admits in *Prince of Networks: Bruno Latour and Metaphysics* [30]. And, as we have previously said, Harman excludes Whitehead because the latter treat individual things as bundles of relations. Finally Harman' object-oriented position is the only one among other Speculative Realists that might be called *both* Heideggerian and Whiteheadian[8].

Contrary to appearances the link between Whitehead and Deleuze is not so evident: the core entity of Whitehead ontology is the 'actual entity' or concrete individual[9] (or 'actants' as Latour calls it to denote human and non-human actors) whereas entities or actors are not the core of reality nor for Deleuze or any other 'virtualists", the first of which Bergson. Bergson avoids breaking reality into discrete states. It is the same for Simondon which presupposes a 'pre-individual' dimension of any reality (Simondon conceived of "pre-individual fields" as the funds making individuation itself possible). And for Manuel DeLanda there is always what it is called a 'space of possibility' [27] deeper than any actualized individual. Deleuze refers to as a *diagram*, a set of universal singularities (they more or less represent *ideal types* in Max Weber terms, but surely not *essences*) that would structure the space of possibilities (or multiplicity[10]) associated with the assemblage. Deleuze defines a diagram as a display of relations of force, or of a distribution of capacities to affect and be affected [9]. To conclude on this point let us say that some thinkers take individual entities as primary (Whitehead, Latour, Harman) whereas some others view them as derivative (Bergson, Simondon, Deleuze[11] or DeLanda).

Whiteheadian 'actual entities' are not a durable substance (they lies behind their accidents, qualities or relations like in tropes 'view promoted by Livet and Nef [8] because they perpetually and instantly perish and be transformed (if they succeed in binding to or prehend other entities) to new actual entities. The same holds for Latour: according to actor–network theory, such actor-networks are potentially transient, existing in a constant making and re-making. This means that relations need to be repeatedly "performed" or the network will dissolve.

[8] Generally, Heideggarians (like Derrida) speak most about the "failures of presence" and not so much about inanimate relations without sentient observers (they are correlationist in some ways) whereas Whitheadian (such as Latour) are relationalist but they are not attracted by the idea of a hidden reality concealed from all presence.

[9] « Actual entity – also termed 'actual occasion' – are the final real things of which the world is made up" ([23], p. 18).

[10] This is why it is not possible to equal multiplicity (Deleuze) and actor-network (Latour) as however many commentators of the two authors do it.

[11] Conversely Deleuze does not speak of the actualization of the virtual in terms of the "things", but in terms of the "event", where an event is inessential, unexpected anomalous, seemingly impossible from the current state of affairs, and therefore capable of opening up the future, making a difference, and changing the world. Every actualization of the virtual is an event [35].

	Whitehead	Deleuze	Latour	Schatzki	Livet & Nef	Harman
Object	yes	no	yes	no	no	yes
Relation	external	external	external	internal	internal	no
Virtuality	no?	yes	no	no?	yes	no
Context	no	no	no	yes	yes	no
Becoming	no?	yes	yes?	no?	no	yes

Figure 2. Some Social Ontologies candidates

On the side of the theories of the practice, in bond with arrangement theory, the things seem more complicated. Frequently practice theorists make a claim in favor of splitting reality in two (articulated) parts: activity (process of production[12]) and some social order which constitute a 'context' for activity: Structure and Agency for Giddens; Field and *Habitus* for Bourdieu; social order and situated practices for Lave, person's identity embedded in a community for Taylor and so on.

Schatzki [37] (p. xi) characterizes thus social life or human coexistence primary by is grounding in something he called "the site of the social". The social site is a specific context of human coexistence, the place where, and as part of which, social life inherently occurs. This site-context is composed of a mesh of orders[13] and practices: "Orders are arrangements of entities (e.g., people, artifacts, things), whereas practices are organized activities" (p. xi). Human coexistence thus transpires as and amid an elaborate, constantly evolving nexus of arranged things and organized activities. By doing that Schatzki argues in favor of *ontology of place* (embedding milieu or medium, closely

[12] Fischbach establishes in a convincing way filiation between Spinoza and Marx around the idea of *ontology of the productive activity* and of its primacy on any other authority. According to Marx and Engels, for individuals, the mode of production is "a definite form of expressing their life, a definite *mode of life* on their part. As individuals express their life, so they are. What they are, therefore, coincides with their production, both with *what* they produce and *how* they produce" [36]. This idea is articulated around a principle common to Spinoza and Marx, by whom this last breaks with the metaphysics of subjectivity suitable for the German idealism: the assertion of the integral inscription of the man in nature. History and Nature are link and form a unity because the natural conditions of the human production are as much its historical conditions. Affirming that, we can measure the cost of the correlationism which brings back, in a typically idealistic epic, the production process of reality to that of its human access. This retreat affects in particular the theory of the practice, in particular in its Wittgensteinian versions, even if the term 'Form of life' (German *Lebensform*) used by Ludwig Wittgenstein connotes the sociological, historical, linguistic, physiological, and behavioral determinants that comprise the matrix within which a given language has meaning.

[13] What made us choose Schatzki among others practice theorists it is that the social order is theorized there as an assemblage. Through Schatzki's work a link between practice and assemblage theories can be found. As he said: "As the master figure organizing this treatise's account of the social, the distinction between arrangements and practices runs throughout the book" (p. xii).

related also to the ideas of Martin Heidegger[14]) and considers that analyzing the social through the concept of site offers several advantages over rival social ontological paths, especially individualist ontologies and also other anti-individualisms, above all 'structuralists" which suffer from a tendency toward hypostatization (fascination for abstract structures).

But the main confrontation is between what he calls himself "theories of arrangements" and "practice theories". The term "arrangement" denotes a group of thinkers who takes arrangements of entities to be the principal compositional feature of social life. If the word "arrangements" does not appear as such, the relevant expressions are instead *apparatus* (or assemblages) (Foucault), *assemblages* (or arrangements) (Deleuze and Guattari) and *actor-networks* (Latour and Callon). Arrangements are "social things organized in configurations, where they hang together, determine one another via their connections, as combined both exert effects on other configurations of things and are transformed through the action of other configurations, and therewith constitute the setting and medium of human action, interaction, and coexistence" (p. xiii).

Almost two main differences cleaves the two sets of categories of social ontology: theories of arrangements are anchored in social *nominalism* which contends that sociality can be explained solely through the properties of and relations among the particular entities that compose social life whereas *contextualism* argues that these matters must be referred to a context, different from these entities, in which the latter exists. By "context" contextualists means a setting or backdrop that envelops and determines phenomena. Examples of contexts are economic systems, social structures, hierarchical distributions of power or capital, webs of meaning, discourses, and social practices. Social Nominalists on the contrary maintain that the character and transformation of arrangements are beholden to nothing but properties of and transactions among the components of arrangements. Individualist ontologies are nominalist in essence whereas 'collectivist' ontologies are not. Theories of arrangements creates new major division because they are nor individualist nor holistic ontologies.

A second ontological issue is nested around humanist and post-humanist confrontation. As humanism assumes numerous forms, humanism consist here in the claim that human agency is both a highest form of agency and have a greatest significance to life on earth. In contrast, post-humanism stresses the causal significance of entities other than humans for social life. It could join the non correlationnist point of view in some ways. Here the boundary between practice theory and theories of arrangement is more porous because thinkers who underline the mediation of intellectual functioning by cultural tools (like Activity Theory or Distributed Cognition Theory but also actor-network theory with its intermediaries and mediators), even those who fall under the thesis of the 'externalism' in the philosophy of mind[15], can be classified sometimes in practice theories camp sometimes in theories of arrangements camp (Whiteheadian process ontology and actor-network theory for sure). The case of Activity Theory is complex because it does not belong to a practice theory current of thought, emphasizes the prominent roles of tools whereas considers that the distinction between subject and object lies in human's agency and, doing so, separated from its theoretical counterparts on the former point, namely Actor-Network Theory. Various post humanists analyses, including theories of arrangements as well as of practice, stress the causal significance of entities other than humans for social life and threaten to dissolve human agency into the actions of nonhuman entities. This is why the moving, the overlap or the 'blurring' of boundary between human and other entities form a radical distinction between social ontologies. Considering these two ontological issues Schatzki condemns the rejection of context and deplores the debunking of the human agency in post humanists works. This criterion meets the question of the correlationnism.

A third issue distinguishes the different conceptions of social and relates to the nature of the order itself. Order is a basic dimension of any domain of entities, if we accept that things tend not to form random aggregates but on the contrary clusters of inter-related determinate stuff. Order is thus the basic disposition of a domain of entities, the way that things are laid out or hang together in that domain. Social order as connection may be figured out by many socio ontological conceptions: order as human 'rational' action, practices (as in practice' version of cultural theory), social structure (holism) or emergence (like in virtual space of possibility modeled by DeLanda from Deleuze). But the main line of divide is between conceptions of order that claims that social life contains perduring and substantive orders and these, more and more numerous, for which organizations and orders are in reality precarious, unstable, and transitory beings. 'Ordering' [38] or 'Organizing' [39, 40] and many other conceptualizations designate by contrast the dynamic processes that contribute to the 'making of' any enterprise (to unique encounter on the street to a complete domain of activity). In this new conceptualization of orders interdependence or coordination are put ahead (whereas regularity or stability): the constraint consisting to be connected to exist subordinates the semantics of the totality and the identity of the entities to coordination as a dynamics and emergent process.

But in the race for various alternatives to well-defined and well-organized enduring wholes, contemporary practices theories, but above all assemblage theories, are on the cutting edge of social ontologies. The issue is that much social ontologies interpret interdependence as regularized ties, interchanges, or reciprocity. And regulated means no arbitrary, which equates order with generic state of affairs. Interdependence connotes also mutual dependence, which is the trademark of the ontological conceptions based on the idea of *internal relations* (that of Wittgenstein, if one believes Descombes [41] in his own version of structural holism. See also [42]): here element is always a part of some sort of whole as if this whole seems to be not fully

[14] By triangulation we can advance that if practice' ontology can be articulated with the 'scaffolding' of Heideggerian - Wittgensteinian ontology (except with regard to the role given to the constraint carried out by the social structure and conflicts in the social reproduction) by contrast it should not be truly compatible with an approach of Whiteheadian - Latourian ontology.

[15] which hold that the mind is not only the result of what is going on inside the nervous system (or the brain) but also of what either occur or exist outside the subject, like Andy Clark or enactivism and embodied cognition.

integrated. Descombes calls the doctrine that he defends, derived from Wittgenstein, "anthropological holism" or "structural holism". It is the view that meaning or thought of any kind inheres in a whole network of practices, institutions, mores, and "forms of life" and form a concept he calls 'objective mind' (in reference to Hegel). This is a holistic position because it maintains that meaning is only given in a totality rather than in a one-to-one relation between a representation and its object. The totality in question is one made up of the human institutions and practices that anthropologists study (illustrated by Yahoo Groups!), which differs from the sort of semantic holism generally discussed by cognitivists (and analytic style philosophers) around action or isolated inference. In holism the concept of an order, of a structure of relations and, above all, rules (in the normative rather that causal sense) is thus crucial but, as we will see further, renew a holistic tradition, that we find wrong, leading from Hegel (or Montesquieu) through Durkheim to functionalist and structural-functionalist period.

So the question of order cleaves two main different ontological assumptions: one which supposes internal relations and, in some ways, an idea of totality; and another which posit the doctrine of external relations and, as Meillassoux calls it, a 'non-totalisable being', a position where structure of the possible as such must necessarily be un-totalizable[16]. Practice theories are usually in favor of the doctrine of internal relations, the work of Livet and Nef too, whereas Deleuze and Guattari and Callon and Latour, although not belonging at all to the same camps according to the nominalist versus contextualist criterion or in virtue of their opposition on the virtual versus non virtual character of assemblages, meet to privilege the doctrines of the external relations.

So we argue in favor of social order as arrangements or assemblages, but not like theories of practice because they are contextualist (activity/context), pro-correlationist and thus non nominalist[17]. For example Schatzki maintain: "What a thing of social life is cannot be fixed. A garden rock, say, can suddenly become a paperweight and at a later moment a weapon (…). In general, both what things are and the state(s) of affairs a given configuration of things constitutes depend on the things involved and their properties in conjunction with how people act toward and understand them" (p. 16). Social nominalism contends that the character and transformation of sociality can be explained solely through the properties of and relations among the particular entities that compose social life (and not by evoking some 'context', different from these entities). Social nominalism is common to Foucault, Deleuze and Guattari, Latour and, as tropes theory, to Livet and Nef (whereas they seem to defend the doctrine of internal relations!).

To conclude, challenging the individual as a basic entity of social ontology, Pierre Livet and Frederic Nef grant this place to the structures, precisely with the structural constraints, of our social activities. The authors decipher then how the interactions proceed, are carried out. And, also challenging the reduction of social reality to our mental constructions, they defend structural realism then: if they admit limiting their analysis to the field of the human activities, they are interested in the methods of the social interaction apart from the experiment that we have some, or how they remain with their investment by the men. They confer a share of virtual thus to them, besides their actuality. The social interactions are the starting point of the ontology of social of Livet and Nef. At the beginning of the analysis, there are not for them initially the individuals, but their relations, which they define as being ways of an element in another. The structure of an interaction becomes a network when the way makes it possible to return to its starting element, when it forms a loop. It happens nevertheless inevitably that breakdowns of network occur. The loop is then not buckled, when an actor is failing or when the environment of the activity prevents, and one does not return to the starting point, at the point of starting of the loop. To these breakdowns of network, Livet and Nef suggest the solution of the substitutability of the activities: In the couple of activities "hunting-gathering", the two activities can replace one the other when one or the other does not succeed. This clarifies their choices to place the 'processes' at the center of social ontology (or ontology of the recursion process). Social reality is thus a phenomenon of potentiality: in an exchange, an activity is virtual whereas the other is current. The ontology of operations attempts to describe networks and virtual processes based on the substitutability of the activities. The substitutability of the activities within an exchange implies internal relations (e.g of a context, as in practice theories) and at the same times the assumption of virtual processes at the core of social operation (like assemblage theories). One can thus say that they choose the anti-correlationnist option and the virtual one, very closed to DeLanda work, but privilege the internal relations, as practice theorists. In doing so, they can't really think a whole as an emergent property from the interactions between componants, as DeLanda do following Deleuze. And by making substantial the contents of the exchanges, they share the object-oriented philosophy, but in Harman's version.

4. CONCLUSION

This paper is a contribution to elicit some strong theoretical claims about conceptual domain ontological options made regarding the empirical social domain. Its purpose is to initiate a novel approach to social ontology around the kind of entities could exist, what could be the relations between them and what could be the form and the properties of social processes. Thus, it investigate some contemporary ontological propositions who share some strong 'air de famille' but also profound differences. We hope that it will help researchers to make their social ontology more coherent.

5. REFERENCES

[1] Appelquist, D., Brickley, D., Carvahlo, M., Iannella, R., Passant, A., Perey, C. and Story, H. 2010. A Standards-

[16] Like Badiou, Meillassoux argues that only those theories that — "ratify the non-All", hence excluding any possible conceivability of a totality, can be defined as ontological, given that being *is* the non-totalisable.

[17] Trope theory in metaphysics could be on a certain plan sympathetic with assemblage theory because it is a version of nominalism. Foucault, who is an eminent representative of the practice theory camps, was absolutely nominalist in his way of studying the forms of power. Power is not a concept, an institution or an abstract structure. "Power must be understood in the first instance as the multiplicity of force relations immanent in the sphere in which they operate and which constitute their own organization" (Foucault [43], *The History of Sexuality*, 1:92). Power thus designates the reticular force relation organization of particulars in the social. Indeed, the social is this reticular organization of particulars.

based, Open and Privacy-aware Social Web. W3C Social Web Incubator Group Report. 6th December 2010 Report. *W3C Incubator Group Report*. Retrieved 6/8/2011.

[2] Giddens, A. 1979. *Central problems in social theory: action, structure, and contradiction in social analysis*. Berkeley, University of California Press.

[3] Coleman, J. 1990. *Foundations of Social Theory*. Cambridge, Mass. Harvard University Press.

[4] Gilbert, M. 1989. *On Social Facts*. Routledge. London.

[5] Searle, J. 1995. *The Construction of Social Reality*. New York. Free Press.

[6] Searle, J. 2010. *Making the Social World: The Structure of Human Civilization*. Oxford University Press.

[7] Smith, B. 1999. Les objets sociaux. *Philosophiques*. Volume 26, numéro 2, automne 1999, 315-347.

[8] Livet P and Nef, F. 2009. *Les êtres sociaux. Processus et virtualité*. Paris, Hermann Éditeurs.

[9] DeLanda, M. 2006. *A New Philosophy of Society: Assemblage Theory and Social Complexity*. London. New York: Continuum.

[10] Meillassoux, Q. 2008. *After Finitude: An Essay On The Necessity Of Contingency*. trans. Ray Brassier. Continuum.

[11] Wittgenstein, L. 1958. *Philosophical Investigations*. New York: Macmillan.

[12] Ortner, S.B. 1984. *Theory in Anthropology since the Sixties. Comparative Studies in Society and History*, Vol. 26, No. 1 (Jan., 1984). Cambridge University Press. 126-166.

[13] Ortner, S. B. 2006. *Anthropology and social theory: culture, power, and the acting subject*. Durham, Duke University Press.

[14] Bourdieu, P. 1977. *Outline of a theory of practice*. Cambridge: New York. Cambridge University Press.

[15] Foucault, M. 1980. *Power/Knowledge*: Selected Interviews and Other Writing, 1972-1977. C. Cordon editor. New York: pantheon Books.

[16] de Certeau, M. 1984. *The Practice of Everyday Life*. trans. Steven Rendall. University of California Press, Berkeley.

[17] Lave, J. 1988. *Cognition in Practice.* Mind, Mathematics and Culture in Everyday Life. Cambridge: Cambridge University Press.

[18] Reckwitz, R. 2002. Toward a Theory of Social Practices. A Development in Culturalist Theorizing. *European Journal of Social Theory*. 5 (2). 245-265.

[19] Lave, J. 2011. *Apprenticeship in Critical Ethnographic Practice*. Chicago: University of Chicago Press.

[20] Turner, S. P. 1994. *The Social Theory of Practices: Tradition, Tacit Knowledge, and Presuppositions*. Chicago, University of Chicago Press.

[21] Schatzki, T. R., K. Knorr-Cetina and Savigny Eike, V. (dir.). 2001. *The Practice Turn in Contemporary Theory*. London. Routledge.

[22] Nicolini, D., Gherardi S. and Yanow D. 2003. *Knowing in Organizations. A Practice-Based Approach*, M.E Sharpe.

[23] Whitehead, A.N. 1929. *Process and Reality: An Essay in Cosmology*. 1979 corrected edition, edited by David Ray Griffin and Donald W. Sherburne, Free Press.

[24] Deleuze, G. and Guattari, F. 1980. *A Thousand Plateaus*. Trans. Brian Massumi, London and New York: Continuum, 2004.

[25] Deleuze, G. and Parnet, C. 2002. *Dialogues*. Columbia University Press.

[26] DeLanda, M. 2002. *Intensive Science and Virtual Philosophy*. London. New York: Continuum.

[27] DeLanda, M. 2011. *Philosophy and Simulation. The Emergence of Synthetic Reason*. London. New York. Continuum.

[28] Latour, B. 2005. *Reassembling the Social- An Introduction to Actor-Network-Theory*. Oxford University Press.

[29] Latour, B. 1988. *The Pasteurization of France*. A. Sheridan and J. Law (trans). Harvard University Press. Cambridge Mass.

[30] Harman, G. 2009. *Prince of Networks: Bruno Latour and Metaphysics*. re.press, Melbourne.

[31] Harman, G. 2011 (a). *The Quadruple Object*. Winchester, UK.

[32] Harman, G. 2011 (b). *Response to Shaviro*. In Bryant, L., Srnicek, N. and Harman, G. (eds.). 2011. The Speculative Turn: Continental Materialism and Realism, re.press.

[33] Bryant, L., Srnicek, N. and Harman, G. (eds.). 2011. *The Speculative Turn: Continental Materialism and Realism*, re.press.

[34] Rescher, N. 2000. *Process philosophy: a survey of basic issues*. Pittsburgh. University of Pittsburgh Press.

[35] Lindsead, S. and Tahnem, T. 2007. Multiplicity, Virtuality and Organization: The Contribution of Gilles Deleuze, *Organization Studies*. 28: 1483. Sage.

[36] Marx, K. and Engels, F. 2001. *The German Ideology Part One*, with Selections from Parts Two and Three, together with Marx's "Introduction to a Critique of Political Economy." New York: International Publishers.

[37] Schatzki, T. 2002. *The Site of the Social: A Philosophical Account of the Constitution of Social Life and Change*. Pennsylvania State Univ Pr.

[38] Law, J. 1992. *Notes on the Theory of the Actor-Network: Ordering, Strategy and Heterogeneity*. *Systems Practice*. 5, 4: 379-93.

[39] Weick, K. 1979. *The Social Psychology of Organizing*. Mc Graw Hill Inc. New York.

[40] Weick, K. 2009. *Making Sense of the Organization: The Impermanent Organization*. John Wiley & Sons Ltd.

[41] Descombes, V. 1996. *Les Institutions du sens*. Paris. Editions de Minuit.

[42] Descombes, V. 2001. *The Mind's Provisions: A Critique of Cognitivism*. Princeton University Press.

[43] Foucault, M. [1976] 1998. *The History of Sexuality* Vol. 1: The Will to Knowledge. London: Penguin.

Filter Bubble and Enframing: On the Self-Affirming Dynamics of Technologies

Andreas Beinsteiner
Ph.D. Student
Institute of Philosophy
Leopold-Franzens-Universität
Innsbruck
andreas.beinsteiner@gmail.com

ABSTRACT

This paper tries to relate the recent concerns about personalized filtering on the internet to Martin Heidegger's philosophy of technology. In "The Filter Bubble", Eli Pariser describes how personalized filtering of online contents may result in a "self loop", amplifying the user's interests and opinions. It will be argued that there are structural similarities between the concept of the filter bubble and Heidegger's concept of technology as enframing. Also the latter addresses a filtered perception of reality which reinforces itself. In both cases, the dynamics under consideration ultimately threaten human freedom. A comparison of filter bubble and enframing might not only produce a deeper understanding of both phenomena, but reveal the discussion of self-affirming dynamics as an essential task for media studies.

Keywords

enframing, filter bubble, Heidegger, technology

1. INTRODUCTION

As the vast amount of data on the Internet is growing faster than ever, filtering becomes a necessity. Internet giants like Facebook or Google have chosen a way that is different from the editorial selection typical of traditional broadcast media. They offer *personalized* filters instead of general ones.

While for a long time this break with the agenda setting of traditional broadcast media was considered an advantage of the Internet, in the new millennium a different a different perspective evolved. In a book released in 2011, Eli Pariser expresses his concerns about this tendency towards personalization on the web. As personalization becomes more and more usual, he argues, we will increasingly become embedded in a filter bubble, in "your own personal, unique universe of information that you live in online." ([10]) The pervasive tendency towards personalization is problematic, as it "moves us very quickly toward a world in which the Internet is showing us what it thinks we want to see, but not necessarily what we need to see." ([10])

The idea is not new: In 2001, Cass Sunstein conceived personalized news as "The Daily We" and wondered if the Internet really was a blessing for democracy (cf. [13]). Also, concerns have been raised for quite a while about a fragmentation of the public sphere, where communication only takes place between people with similar interests and attitudes.

My claim in this paper is that the filter bubble may serve as a model that illustrates a more general concept about the self-affirming dynamics of our technologies: Martin Heidegger's concept of the enframing. Relating these two ideas may result in mutual benefits: it might help to establish a better understanding of Heidegger's notoriously difficult and notoriously misunderstood concerns, and in turn the kind of problem that Eli Pariser calls attention to may be grasped more precisely in Heidegger's terms. Ultimately the paper suggests developing the concept of a local enframing as a critical tool for media studies.

2. FILTER BUBBLE

Facebook and Google were the places where Pariser first became aware of the effects of personalization. "I noticed one day that the conservatives had disappeared from my Facebook feed", he tells us. "And what it turned out was going on was that Facebook was looking at which links I clicked on, and it was noticing that, actually, I was clicking more on my liberal friends' links that on my conservative friends' links. And without consulting me about it, it had edited them out." ([10])

The same kind of editing, Pariser found out, also happened on Google: He asked two friends to search for "Egypt" on Google. The results were drastically different: "Daniel didn't get anything about the protests in Egypt at al in his first page of Google results. Scott's results were full of them. And this was the big story of the day at that time. That's how different these results are becoming." ([10])

Personalization is used at a lot of other places too: On online dating platforms, obviously, but also more and more on news portals. Why is Pariser worried about this development? He sees a number of problematic effects that occur with the rise of filtering.

One serious consequence for democracy is the decline of the public sphere: "In the filter bubble, the public sphere – the realm in which common problems are identified and addressed – is just less relevant." ([11], p. 148) Another one is the "friendly world syndrome": "[S]ome important public problems will disappear. Few people seek out information about homelessness, or share it, for that matter. In general, dry, complex, slow moving problems – a lot of the truly significant issues – won't make the cut." ([11], p. 150f) This relates to another issue: "[I]nstead of a balanced information diet, you can end up surrounded by information junk food." ([10])

However, at the centre of all these tendencies there is one effect that Pariser calls "the you loop:" "The filter bubble tends to dramatically amplify confirmation bias – in a way, it's designed to. Consuming information that conforms to our ideas of the world is easy and pleasurable; consuming information that challenges us to think in new ways or question our assumptions is frustrating and difficult." ([11], p. 88) Personalized filtering directs us towards doing the former: „[T]he filter bubble isn't tuned for a diversity of ideas or of people. It's not designed to introduce us to new cultures. As a result, living inside it, we may

miss some of the mental flexibility and openness that contact with difference creates." ([11], p. 101) This is not only a danger for democracy, but also for freedom. For freedom, Pariser explains, cannot be reduced to being able to do what you want. First you need to know what is possible to do. (cf. [11], p. 112) "When you enter a filter bubble, you're letting the companies that construct it choose which options you're aware of. You may think you are the captain of your own destiny, but personalization can lead you down a road to a kind of informational determinism in which what you've clicked on in the past determines what you see next – a Web history you're doomed to repeat. You can get stuck in a static, ever-narrowing vision of yourself – an endless you-loop." ([11], p. 16)

Also Cass Sunstein perceived filtering as a threat to democracy and freedom.[1] An important difference, however, is that Sunstein was concerned with personalization that the user consciously chooses. This does not hold in the filter bubble: "When you turn on Fox News or read The Nation, you're making a decision about what kind of filter to use to make sense of the world. It's an active process, and like putting on a pair of tinted glasses, you can guess how the editor's leaning shapes your perception. You don't make the same kind of choice with personalized filters. They come to you – and because they drive up profits for the Web site that uses them, they'll become harder and harder to avoid." ([11], p. 10) So for Pariser maybe the most dangerous thing about filter bubbles is that they are not aware of them: "In fact, from within the bubble, it's nearly impossible to see how biased it is." ([11], p. 10) Accordingly, the mission of his book, and the first step towards solving the problem, is to render the filter bubble visible (cf. [11], p. 20). A second step would be to think about how serendipity happens and how it could be promoted by software design decisions (cf. [11], p. 235f).

My claim in what follows will be that the filter bubble can serve as a model to understand a more generic concept about the self-amplifying dynamics of technologies: Martin Heidegger's concept of technology as *enframing*.

3. ENFRAMING

Martin Heidegger's esoteric and idiosyncratic terminology has given rise to a lot of misunderstandings. As often, building bridges between different kinds of vocabularies might help to clarify things. I will try to do this by relating the dynamics Eli Pariser describes to the ones Heidegger describes.

If for Pariser it is the enormous amount of data online that requires filtering, Heidegger's concern is the inexhaustible richness of *Being* getting filtered. Being is no mythic or divine entity but simply refers to the meaning of the word "to be", to our understanding of what it means that something "is". So Being obviously is strongly intertwined with language.

The inexhaustible richness of Being is not a kind of mythological postulate, but a simple consequence of the historicity of Being. As Heidegger tries to show, every culture and epoch had its own understanding of Being. This understanding changes, as language changes. And as we have no idea about how language might develop, there is no basis for determining a definite set of *possible* understandings of Being.

As a consequence, every historical culture has its own *clearing of Being*, which is at the same time *concealment* and *unconcealment*. While a vast majority of possible understandings of reality remain hidden for us, a certain understanding unfolds. Thus with every clearing of Being, only certain few aspects of reality become accessible for us.

This means that the respective clearing defines our possibilities in thinking and acting, and, more fundamentally, our possible horizons of meaning. Richard Rorty puts it this way: "For Heidegger – early and late – what one is is the practices one engages in, and especially the language, the final vocabulary, one uses. For that vocabulary determines what one can take as a project." ([12], p. 109)[2] The clearing of Being thus provides a refined concept for discussing a matter that also Pariser is concerned with: "Not knowing that it is possible to be an astronaut is just as much a prohibition against becoming one as knowing and being barred from doing so." ([11], p. 112f.) The matter at stake is freedom.

If the clearing of Being is historical, it can change. With Heidegger, we can grasp freedom precisely as the *mutability* of the conceptual framework that mediates our access to reality. Freedom relies on what I want to call *hermeneutic oscillation*: on a condition where various modes of unconcealment are suspending and balancing each other.[3]

We can consider the clearing of Being as a filtering of the inexhaustible richness of Being: "Beings can be as beings only if they stand within and stand out within what is cleared in this clearing." ([7], p. 178) Freedom thus requires that this filtering must not become static: It must not always be the same aspects of reality that get filtered out or that make it through the filter. Only then new aspects of reality may appear and provide us with new possibilities of thinking and action.

Heidegger's history of Being could thus be rephrased as a history of filters on the possible meanings of Being. In the modern age, or, as Heidegger puts it, "the Age of the World Picture", "an essential decision takes place regarding what is, in its entirety." ([8], p. 130) Man is understood as the *subject* and all entities become *objects*: "Man becomes that being upon which all that is, is grounded as regards the manner of its Being and its truth. Man becomes the relational center of that which is as such." ([8], p. 128) As man is placed as the relational center of everything that is, the world becomes a picture, a representation for him.

"Here to represent [vor-stellen] means to bring what is present at hand [das Vorhandene] before oneself as something

[1] "Unanticipated encounters, involving unfamiliar and even irritating topics and points of view, are central to democracy and to freedom itself." ([13])

[2] As we will see in the course of the following considerations, the clearing of being is not only constituted by language, but also by technologies and technical artefacts. Both aspects have been addressed frequently in Heidegger's writings.

[3] For Heidegger's concept of freedom, cf. "On the Essence of Truth, p. 115-138 in [7]. Understood in this way, freedom is not something that man possesses as a property. "At best, the converse holds: freedom, ek-sistent, disclosive Da-sein possesses man – so originally, that only it secures for humanity that distictive relatedness to being as a whole which first founds all history." ([7], p. 127)

standing over against, to relate it to oneself, to the one representing it, and to force it back into this relationship to oneself as the normative realm." ([8], p. 131) This means that everything that is, *is*, insofar it can be related to man. With other words, everything that cannot be related to man is filtered from the clearing of Being: all aspects of reality that are not relevant to man cease to exist.[4]

My suggestion is that the age of the world picture, as a reconfiguration of the ontological sphere, structurally corresponds to the introduction of personalized filtering on the internet: Everything that cannot be related to oneself as relevant in some way is filtered out of existence.

For Heidegger, the modern reconfiguration of the clearing of Being has serious ethical consequences, as it implies that in everything man does, he is only concerned with aspects of reality that in some way relate to himself. However, this constellation is radicalized with the advent of modern technology.

Heidegger calls the *Wesen* of technology the enframing [*Ge-stell*]. While Wesen usually is translated as essence, I suggest that *ontological dynamics* is a more appropriate translation. For in Heidegger the notion does not refer to any supposed nature of things, but to the way they relate to changes in the clearing of Being.[5] Technology, according to Heidegger, is not merely a means. "Technology is a mode of revealing. The dynamics of technology are situated in the realm where revealing and unconcealment take place, where aletheia, truth, happens." ([7], p. 319[6])

What are those ontological effects of technology that Heidegger describes as Ge-stell? Everything is revealed only as *standing-reserve* [*Bestand*], things may only enter the clearing of Being to the extent that they can be conceptualized as an instrument or challenged as a resource. Observe that the Ge-stell mode of revealing is narrower than the world-picture mode: "Whatever stands by in the sense of standing-reserve no longer stands against us as object." ([7], p. 322) Now what in some way relates to man may no more pass through the filter but only that which is useful for our purposes. This means that our possibilities of being in the world become more narrow too:[7] Heidegger contrasts the river Rhine, technologically perceived as an energy supplier or as a tourist attraction, with the Rhine as it appears in the poetry of Hölderlin (cf.[7], p. 321).

Technology filters reality in a way so that we perceive only the aspects of reality where it is successful.[8] And the more we perceive technology as successful, the more it will reinforce not only its own take on reality, but also the corresponding horizons of meaning that drive our activities. "Man clings to what is readily available and controllable […], concealing as a fundamental occurrence has sunk into forgottenness." ([7], p. 132f) In a similar way, Pariser states that the filter bubble transforms "known unknowns into unknown unknowns". ([11], p. 106) We can explore this structural resemblance further: "Left to their own devices, personalization filters serve as a kind of invisible autopropaganda, indoctrinating us with our own ideas, amplifying our desire for things that are familiar and leaving us oblivious to the dangers lurking in the dark territory of the unknown." ([11], p. 15) As everything is filtered that is mysterious or does not fit into the established conceptualizations, those conceptualizations become static. "By disavowing itself in and for forgottenness, the mystery leaves historical man in the sphere of what is readily available to him […]," states Heidegger. And Pariser stresses that "[i]f personalization is too accute, it could prevent us from coming into contact with the mind-blowing, perception-shattering, experiences and ideas that change how we think about the world and ourselves." ([11], p. 15)

Both in the filter bubble and in enframing, man is stuck in a certain conceptualization of reality. When Heidegger says that thus the essence [Wesen] of man is threatened by technology (cf. [7], p. 333), this does not involve any essentialist claims about the nature of man. On the contrary, it means that the ontological dynamics of man have come to stagnate; that the indefinite possibilities of what man might be have been narrowed down to one single understanding of man that is amplified and reinforced by the relational system of our technologies. In the same way, "the economics of personalization," according to Pariser, "push toward a static conception of personhood." ([11], p. 216)

---

[4] As the man of the modern age, according to Heidegger, is metaphysically conceived as *animal rationale*, there are first and foremost two modes of that relation: either things serve as an *experience* [*Erlebnis*] for man as an animal, or things can be *measured* scientifically by rational man.

[5] Although Heidegger dedicates several pages to explaining his reinterpretation of the term (cf. e.g. [7], p. 334ff), many interpreters still hold on to the traditional notion of Wesen.

[6] Translation modified. German original: „Die Technik west in dem Bereich, wo Entbergen und Unverborgenheit, wo aletheia, wo Wahrheit geschieht." ([6], p. 17) This statement has to be read carefully: The realm, where the dynamics of technology, understood in the Heideggerian sense as enframing, are situated, is the clearing of Being. In many of his writings, Heidegger indicates that this clearing is not only constituted by language, but also by artifacts, tools and machines (cf. e.g. the *tool analysis* in "Being and Time"). So there is always a technological aspect in the clearing of Being. This aspect, however, is to be distinguished from enframing as a certain tendency in the dynamics of the clearing of Being in the age of technology. For any attempt to estimate Heidegger's relevance for media studies, it is essential to clarify the exact relation between technological artifacts and the tendency of enframing: *What kinds* of artifacts and infrastructures do promote enframing, and *why*?

[7] "The only thing that is ever questionable is how we can measure and fathom and exploit the world as quickly as possible, as securely as possible, as completely as possible […]." ([2], p. 41f.)

[8] Quantification is an essential feature in the filtering that renders technological access to reality successful: "Calculation refuses to let anything appear except what is countable. Everything is only whatever it counts. […] Only because number can be infinitely multiplied, irrespective of whether this occurs in the direction of the large or the small, can the consuming dynamics of calculation hide behind its products and lend to calculative thinking the semblance of productivity - whereas already in its anticipatory grasping, and not primarily in its subsequent results, such thinking lets all beings count only in the form of what can be set at our disposal and consumed.." ([3], p 235, translation modified)

This section tried to make transparent that Heidegger's philosophy of technology does not address any supposed opposition of nature and technology, but a self-amplifying dynamics that structurally resembles the one described by Eli Pariser's filter bubble. One thing we can learn from these similarities is that there are actually multiple ways of drawing on Heidegger for a critical enquiry into today's media environments. The more traditional way would be analyze if and how media and technological infrastructures contribute to enframing by making everything available as a standing reserve. However, for those who do not accept Heidegger's narrative about the totality of technological access to the world in our age, there exists another way of making use of Heidegger's considerations. Since enframing, like the filter bubble, is about self-amplifying dynamics, the concept can also be employed without any claims of totality, to identify *local enframings*:[9] conceptual frameworks that reinforce themselves, horizons of meaning that we have become stuck in without being aware of it. In this approach, *concrete* media or technological infrastructures could be analyzed with respect to their ontological dynamics: Do they promote hermeneutic oscillation or do they establish local enframings? Do they allow for a mutability of concepts, or do they reinforce established understandings? The first crucial step in destabilizing local enframings, however, might be to *realize* that we always are exposed to a clearing of Being that is constituted by our language and our technologies and that is in danger of becoming static.

4. THE SELF-AFFIRMING DYNAMICS OF TECHNOLOGIES

According to Heidegger, the lock-in in one clearing of Being is particularly strong because man is not aware of the filtering that is at work in this clearing: "Man stands so decisively in subservience to on the challenging-forth of enframing that he does not grasp enframing as a claim, that he fails to see himself as the one spoken to, and hence also fails in every way to hear in what respect he ek-sists, in terms of his ontological dynamics [Wesen], in a realm where he is addressed […]." ([7], p. 332, translation modified)

After Heidegger had conceived the revealing in the mode of enframing as the *supreme danger*, his text takes an irritating turn. He refers to a verse of Hölderlin to declare that "where danger is, grows [t]he saving power also." ([7], p. 333)

This might seem arbitrary, but Heidegger explains: "The danger itself, if it is *as* the danger, is the saving power." ([8], p. 41) If the danger becomes perceived explicitly as the danger, this might free us from the lock-in in enframing: "[W]hen we once open ourselves expressly to the *ontological dynamics* [Wesen] of technology we find ourselves unexpectedly taken into a freeing claim." ([7], p. 331, translation modified) Heidegger seems to hope that, as the ontological dynamics of technology become more intense, they might also become visible as such: as a selective filtering that amplifies established concepts and horizons of meaning.

Self-affirming dynamics are not exclusive to technology but denote a danger that *always* threatens man. Being exposed to the potential infinity of possible ways of conceptualizing the world, man tends to hold on to those kinds of conceptualizations that he already is familiar with: "As ek-sistent, Dasein is insistent." ([7], p. 132) Also for Pariser, the consumption of news that confirm one's own belief existed before the filter bubble. "And while this phenomenon has always been true, the filter bubble automates it. In the bubble, the proportion of content that validates what you know goes way up."([11], p. 89)

I want to suggest that the *automation* of the phenomenon might be a crucial point. As the self-affirming dynamics in questions have been objectified into software by several different internet platforms who offer personalization, and as Pariser has written a book about it, the problem has become explicit. In a way, the danger now is unconcealed *as* the danger. Pariser's aim was to render the filter bubble visible, just like Heidegger's concern was whether enframing would reveal itself as such.

The structural similarities of filter bubble and enframing indicate that a useful notion of a local enframing can be developed. However, one has to restrain from a premature identification of filter bubble and enframing. Instead, the differences of the two concepts have to be clarified. Here only a few of these differences are exposed in order to raise some productive questions.

I. One difference is that the filter bubble seems to be an *epistemological* problem, while the enframing is an *ontological* one. The filter bubble defines what we are able to find out about, while the clearing of Being ultimately defines what *is*. This is the case, because the filter bubble is not our only access to reality. We also find out about things when we are not online. But, in contrast, there is nothing outside of the clearing of Being.

This difference, however, might blur, as we spend more and more of our lives online and as the internet begins to colonize our offline world with the development of augmented reality. If ultimately, as Pariser describes (cf. [11], p. 207ff), our whole lives might be absorbed by the filter bubble, would those filters thus obtain the ontological totality that Heidegger envisioned? This might depend upon whether interactions with other individuals in the social sphere might allow us to break through the filters. We are thus lead to another important difference.

II. While, according to Pariser, every individual human being is enclosed in its own filter bubble, Heidegger became less and less concerned with individuals in the course of his philosophical career. After the individualistic "Being and Time", he came to be more occupied with the fate of the Germans as a "historic people". After World War two, when his philosophy of technology took shape, he was interested in mankind as such, since he supposed that with technology, European thinking had pervaded the whole globe. Every individual is enclosed in its own filter bubble, but the whole mankind is enclosed in enframing. So while social interaction and communication might crash our individual bubbles, Heidegger is interested in the basic understandings that we all already take for granted and which thus cannot be shaken that easily by communication. In particular this is the case for understandings which are entailed by the communication

[9] The notion of a local enframing is choosen in order to discard two aspects of totality connected to Heidegger's original concept: 1. the *planetary* dimension of enframing which pervades all of the contemporary world and 2. the *totality* of the specific interpretation of reality that Heidegger links to the self-affirming dynamics of technology. As we can learn from Pariser, such dynamics may be *restricted* to small groups or even individuals, and they are not necessarily linked to *this specific* understanding of reality. The point here is not to qualify the scope of Heidegger's cultural diagnostics, but to enhance the applicability of the concept for analyzing self-affirming dynamics in a variety of concrete technological settings.

infrastructures themselves.[10] Moreover, it is the case for understandings that influenced the design of our communication technologies.

This raises also the question about the relation of individual filter bubbles and the ones that pervade the whole society. Filtering algorithms are developed and programmed on the basis of certain established understandigs and horizons of meaning. As Heidegger mentions, the "functionaries" for "making public civilized opinion" are "at once driver and driven" ([4], p. 212), they constitute understandings and are constituted by them. If one tries to follow Pariser's suggestion and looks for ways to design and implement serendipity (cf. [11], p. 235f), one has to be aware: Also programming decisions tend to be a result of a filtered perception of reality.

III. There is a certain ambivalence within Pariser's grasp of the problem. Although he generally seems to be concerned about a loop that reinforces ones attitudes and interests (as described in Section 2), there are also some passages where he seems to be worried that the *authentic self* of the user could become *manipulated* by the filters: "You become trapped in a you loop, and if your identity is misrepresented, strange patterns begin to emerge, like reverb from an amplifier." ([11], p. 125) If the problem was only that of a misrepresentation of an authentic self, then building Popperian falsification strategies into the filtering algorithms (cf. [11], p.132ff) might really help. This understanding of the problem, however, drops the insight, how deeply we are shaped by our language and our technologies. [11]

As he became aware of how fundamentally man was constituted by the clearing of Being, Heidegger stopped using his early notion of authenticity. "For there is no such thing as a man who is as a man singly and solely by his own virtue." ([7], p. 337, translation modified) It is worth noting that Heidegger's "Question Concerning Technology" contains an answer to Werner Heisenberg. Heisenberg had described the technological age as a condition where man always and everywhere only encounters himself (cf. "Das Naturbild der heutigen Physik", pp. 109-127 in [9]). Heidegger objects that man "ek-sists, in terms of his ontological dynamics [Wesen], in a realm where he is addressed, so that he *can never* encounter only himself." ([7], p. 332, translation modified, emphasis in the original text) The realm where he is addressed is the historical clearing of Being which is constituted by language and technological artefacts. In this respect Heidegger agrees with many contemporary theorists of media and technology: There is no authentic pretechnological self. However, such a notion of authenticity is not needed as normative concept for critique, since the Heideggerian understanding of freedom as hermeneutic oscillation provides an alternative normative perspective. From this perspective, the identification of self-amplifying tendencies in our technologies – which is, though on different levels, the aim of both Heidegger and Pariser – assumes an essential role in any critical inquiry into our evolving online media environments.

[10] How communication technologies interfere with language became an essential question in several of Heidegger's late writings. E.g., in [1], Heidegger addresses the univocity (i.e. the suppression of hermeneutic oscillation) that language needs to assume in order to become suitable for automated data processing. This is an issue that continues to be relevant as the *semantic web* emerges. It might be instructive to discuss *ontology engineering* in the context of Heidegger's critique of metaphysics.

[11] Maybe Pariser's occasional worries about the manipulation of an authentic self express discomfort about the fact that this sphere, where man is addressed and constituted, is, to an increasing extent, organized according to the interests of private and profit-oriented corporations. Pariser calls this "the commercialization of everything – even of our sensory apparatus itself." ([11], p. 215) Doubtlessly, the static self produced by filter bubbles is useful for profit-oriented enterprises as it makes the behaviour of consumers computable. For Heidegger, however, the desire to calculate human behaviour is not just an effect of capitalism but rather of the cybernetic paradigm in the age of technology.

5. ACKNOWLEDGMENTS

This work was supported by a stipend of the Vizerektorat für Forschung, Leopold-Franzens-Universität Innsbruck.

6. REFERENCES

[1] Heidegger, Martin. 1989. *Überlieferte Sprache und technische Sprache.* Erker: St. Gallen.

[2] Heidegger, Martin. 1996. *Hölderlin's Hymn "The Ister"*. Indiana University Press, Bloomington.

[3] Heidegger, Martin. 1998. *Pathmarks.* Cambridge University Press, Frankfurt am Main.

[4] Heidegger, Martin. 2001. *Poetry, Language, Thought.* HarperCollins, New York.

[5] Heidegger, Martin. 2002. *Die Technik und die Kehre*. Klett-Cotta, Stuttgart.

[6] Heidegger, Martin. 2004. *Vorträge und Aufsätze.* Klett-Cotta, Stuttgart.

[7] Heidegger, Martin. 2008. *Basic Writings*. HarperCollins, New York.

[8] Heidegger, Martin. 2008. *The Question Concerning Technology and Other Essays*. Harper & Row, New York.

[9] Heisenberg, Werner. 1971. *Schritte über Grenzen.* Piper, München.

[10] Pariser, Eli. 2011. Beware online filter bubbles. TED talk. http://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles.html.

[11] Pariser, Eli. 2011. *The Filter Bubble*. Viking, London.

[12] Rorty, Richard. *Contingency, Irony, and Solidarity.* Cambridge Unversity Press, New York.

[13] Sunstein, Cass. 2001. The Daily We. Boston Review. http://bostonreview.net/BR26.3/sunstein.php

Where do ‘ontologies’ come from? Seeking for the missing link

Aurélien Bénel
Troyes University of Technology
http://benel.tech-cico.fr/

Abstract – One of the possible matters for discussion between Web architects and philosophers relies in the use of the term ‘ontology’ by the former. Whether many computer scientists declare that their ‘ontologies’ have nothing to do with the philosophical concept, we must note the analogy between their positions and the positions of the Logical Positivism in the 1930s. However, drawing a guaranteed lineage is extremely difficult. Indeed, in computer science papers, bibliographical references to ontologies usually lead to 1991 only. Hence, this paper is an ‘inquiry’ in search of a chain of descent from the 1930s to 1991.

Keywords – Epistemology, Terminology, Artificial Intelligence, Semantic Web.

Ontologies : what for?

The link between the Web and philosophy is not straightforward. However, in 2001, a wide audience discovered in *Scientific American* that:

> A program that wants to compare or combine information across (...) two databases has to know that (...) two terms are being used to mean the same thing. Ideally, the program must have a way to discover such common meanings for whatever databases it encounters. A solution to this problem is provided by the third basic component of the Semantic Web, collections of information called ontologies.
>
> (Berners-Lee *et al*., 2001)

Hence, faced to a mundane computing problem, Web architects summoned a concept – or at least a term – from one of the most ancient and arduous domains of western philosophy:

> In philosophy, an ontology is a theory about the nature of existence, of what types of things exist; ontology as a discipline studies such theories. Artificial intelligence and Web researchers have co-opted the term for their own jargon, and for them an ontology is a document or file that formally defines the relations among terms. The most typical kind of ontology for the Web has a taxonomy and a set of inference rules.
>
> (Berners-Lee *et al*., 2001)

Critics

The mention of taxonomies briskly worried people who regularly use thesauri and classification systems in libraries. According to Clay Shirky, for example, ontologies have no reasons to age better than 'Marxism-Leninism' category in soviet libraries, or to be less ethnocentric than 'History' in the Library of Congress headings. Without denying the importance of such systems, he thinks that they should be confined to domains in which what is described is stable and restricted, in which categories are well formalised with clear edges (Shirky, 2005).

Years before, while some advocated for *reusable* (Gruber, 1991) or *portable* ontologies (Gruber, 1993) in knowledge bases, our French community of knowledge engineering expressed their reluctance to the alleged universality of ontologies:

> The task strongly influence the building of the ontology which, henceforth, cannot be portable nor universal. Moreover, this advocates in favor of a non-logical but rather constructivist vision of knowledge. (...). As any knowledge, ontologies are

interpreted by a human expert, depending on the idea he has about the task attributed to the system.

Translated from (Charlet, Bachimont *et al.*, 1996)

Models are not problematic by themselves but by the truth status ones give to them, and the last century precisely brought a drastic change in the definition of truth (Léonhardt, 2008). While truth had been defined since Aristotle as the *correspondence to the World*, its recent redefinition led to modern mathematics and modern sciences (see Figure 1).

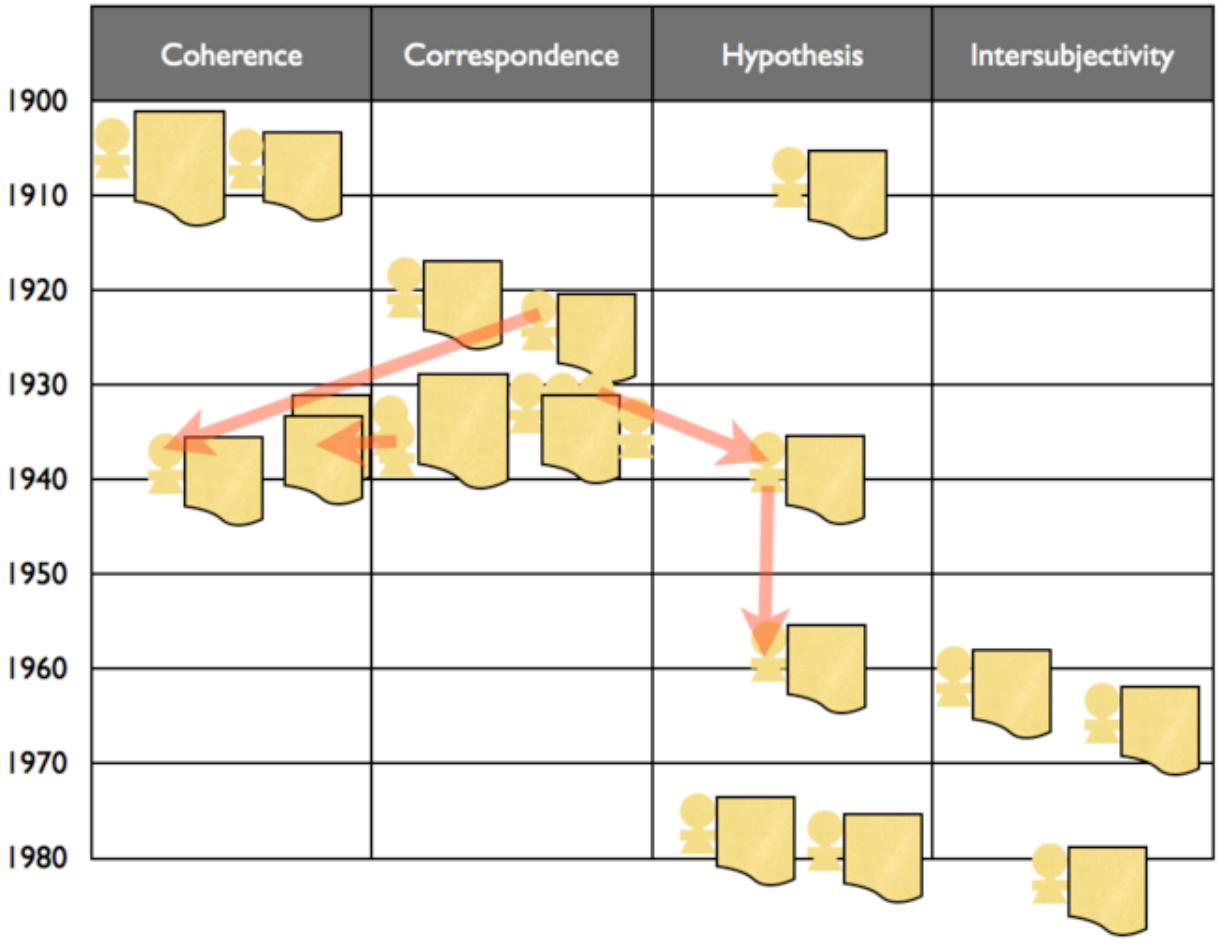

Figure 1. On the definitions of truth.

Hilbert's program (1900-1930), by refounding geometry with formal theorems built on conventional axioms, publicized the idea that truth in mathematics was a matter of *inner coherence*. This revolution was though greatly prepared by the advent of non-euclidian

geometries, new imaginary ‘worlds’ that denied the idea that physics and mathematics shared the same object.

From Hilbert's and Poincaré's mathematics, Schlick (1917) inferred that in order to keep *truth as correspondence* in mathematized physics, axioms had to be replaced with experimental results. In Kant's terms: synthetic statements could only be *a posteriori.* Dedicated to Schlick, *The Scientific Conception of the World – The Vienna Circle* (1929) followed this trail through by founding “Logical Positivism” at the crossroad of empirism and logicism. Paradoxically, the critical impact of the Vienna Circle on modern epistemology was through the people on its margins.

Firstly, Popper (1934), by studying how experimental results and logic could be used together to build scientific knowledge, discovered that Schlick's method was indeed an induction, a logical fallacy condemned since Aristotle. Instead, he then proposed to build scientific method on *modus tollens* deduction: “if it is false for several, it cannot be true for all”. A particular experiment brings knowledge when it refutes a general theory. Therefore, while the falseness of a scientific theory can be certain, its truth is always *hypothetical.* Contrary to similar works (Duhem, 1906), Popper's ones had a major impact on modern science epistemology.

Secondly, Wittgenstein (1936), who was one of the early inspirers (1921) of the Vienna Circle, publicly denied his prior works. By introducing concepts like “Language games”, he clearly gave up the *correspondence* definition of truth for the *coherence* one.

Thirdly, Kuhn (1962), in the Vienna Circle series (*International Encyclopedia of Unified Science*), published a study about scientific revolutions in history. This study, by highlighting the social nature of scientific truth, reminded earlier Marxists works about the prominent role of communities in science, and opened a new field in the sociology of science.

This short historical incursion in the definitions of truth can help us analyse the status given by computer scientists to their ontologies. As an example, Nicola Guarino clearly advocates for “truth as coherence” when he claims:

> The general perspective I have in mind is that of Formal Ontology, which can be included as the theory of formal distinctions between the elements of a domain, independently of their actual reality.
>
> (Guarino, 1997)

In most of computer science works, the lack of reflection about creating, testing and revising ontologies seems to anachronically match the definition of truth advocated by Schlick (*truth as correspondence to the World*), with absolutely no place for knowledge construction and refutation.

On the contrary, some of us think that users of our systems should have the hypothetical and intersubjective value of truth in mind. Hence knowledge engineering should provide digital spaces for debate between contradictory user-generated viewpoints (Bénel *et al.*, 2001; Cahier & Zacklad,2001).

A missing link

We supposed that orthodox ontologists' posture could be explained by a simple 'residual' positivism in the scientific community rather than by a true filiation to Logical Positivism. But then, how could we explain the use of the term `ontology'?

In the English community of knowledge modeling, one of the few authors who refer to philosophical readings is John F. Sowa (Sowa, 1992). However, when he used the word `ontology' (Sowa, 2001), it was to criticize fiercely an artificial intelligence that would not take into account the failure of Logical Positivism.

Because it does not seem to be a direct filiation, we will study in the next sections two hypothetical trajectories of the idea and term of 'ontology'.

In search of heirs

The trail of ontology is easy to follow from Greek philosophy to Logical Positivism, but it fades after the collapse of the Vienna Circle. Owing to Monique Slodzian's works, we know that, at this time in Vienna, a certain Eugen Wüster saw himself as the true heir of the Circle.

Contrary to the original members, Wüster was neither a physicist nor a mathematician, nor a logician, nor a philosopher. He was an entrepreneur who saw in the scientific program of the Circle the opportunity to solve the communication problem between engineers speaking different languages. To address this problem, he defined the "General Theory of Terminology". From language he kept neither verbs nor syntax, but only 'terms'. These terms are structured into what he called himself "an ontology" (Slodzian, 2006).

It is worth noting that Wüster's thesis was entitled "International Language Standardization in Technology" (1931) and that he was indeed at the origin of one of the ISO commissions. So, between Vienna Circle's program and Wüster's one, the goal had significantly changed: while the former did science, the latter did engineering, while the former aimed at describing Nature, the latter described artifacts. In the end, the Ontology became a nomenclature, and the term a purely conventional symbol (Slodzian, 2006).

In fact, Wüster's program was not very far from what François Rastier (Rastier, 2010) criticizes in "Web Science", not far either of what Tim Berners-Lee himself states in his interview with Harry Halpin and Alexandre Monnin:

> When we design a protocol, we're actually creating... we get the chance to actually define the way a new world works. (...) When you create a protocol, you

> get the right to play God, to define what words mean. (...) People (...) have to join in, (...) with agreeing.
>
> Tim Berners-Lee in (Halpin & Monnin, 2010)

In search of ancestors

Another way to draw a lineage from Logical Positivism to the Semantic Web, could be to track down references in papers recursively. But here again, the trail fades. Bibliographical references to ontologies in computer science papers all leads to a short paper by Thomas Gruber (1991). This article itself contains only six references, the oldest ones being from the previous year.

Facing such a dead end, we are reduced to searching for quantitative clues in bibliographic databases (see Figure 2).

Decades	Papers counts	Most cited authors
1930-1939	$3{,}8*10^2$	Coomaraswamy, Ginzburg, Somerville
1940-1949	$6{,}0*10^2$	Cerf, Quine, Cellars
1950-1959	$1{,}5*10^3$	Quine, Cartwright, Wells
1960-1969	$3{,}1*10^3$	Sartre, Bazin, Sommers
1970-1979	$5{,}8*10^3$	Cavell, G'ivon, Hellman
1980-1989	$1{,}4*10^4$	Moens, Cracraft, Horgan
1990-1999	$6{,}6*10^4$	Gruber, Guarino, Ushold
2000-2009	$3{,}1*10^5$	Euzenat, Ashburner, McGuinness

Figure 2. Search in Google Scholar for papers containing the word 'ontology'.

The first straightforward observation is the geometric growth of the use of the term 'ontology'. However this is probably biased by the lack of representativity of digitized contents depending on their age, and mostly by the explosion of scientific papers numbers in the 20th century. What is greatly more interesting is the evolution in the trends witnessed by the names of the most cited authors.

In the 1930s, at the time of the Vienna Circle, 'ontology' is still mainly used in religious studies and philosophy of science. The effects of the Vienna Circle show up in the 1940s and 1950s with the advent of Quine in the most cited authors. In the 1960s, analytical philosophy is overshadowed by phenomenology and reflections on art. It returns to the fore in the 1970s, along with a Viennese-inspired linguistics. This trend seems reinforced in the 1980s in the form of a 'computational' linguistics. Finally, we find Thomas Gruber's and Nicola Guarino's knowledge representations in the 1990s and the Semantic Web in the 2000s.

Of course the coarse-grained results of such a quantitative analysis are not quite satisfactory. However, we can note that, even if the lineages are still blurry, there is a real chronological continuity in the use of the term 'ontology' from Logical Positivism to computer science. Moreover, it is noteworthy that one of the key links in this continuity seems to be Quine.

Further researches on those who deal with ontologies and conjure up Quine lead us to John McCarthy in 1980, the same who introduced the concept of Artificial Intelligence at the famous Dartmouth Conference in 1956. According to him, builders of logic-based intelligent systems must first "list everything that exists, building an ontology of our world" (Smith & Welty, 2001).

By following the same trail, one discovers that the first to follow this advice was Patrick Hayes in 1985. His ontology was for a "naive physics". The word 'naive' was used here not in the sense of a simulation of human reasoning in everyday life, but in the sense of grasping the world pre-theoretically and reasoning about it formally (Smith & Casati, 1993). Nowadays, Patrick Hayes is involved in developing RDF-core, SPARQL and OWL, three core building blocks of the Semantic Web.

A loose link?

Thus emerges a direct filiation through Quine and Artificial Intelligence between logical positivism and the ontologies of the Semantic Web. However, we must admit that Quine's views are rather different from Schlick's. Moreover, matching all of the views of Quine with one definition of truth would be difficult as he denied the distinction between analytic and synthetic statements, and was even opposed to the idea of a normative epistemology.

To go further, we will focus on Quine's reflections *on what there is* (1948) and how they could have been of interest to McCarthy and his logic-based intelligent systems. In this very paper, he openly dismisses the idea that existence would be discovered or invented. According to him, reference is not a matter of *names* but of *pronouns*("bound variables" in formal logic).In other words, the referent is internal to language. As peculiar as this position can be for an 'ontology', it had tremendous advantages for the founder of Artificial Intelligence:

- explained in formal logic terms and with continuous references to Frege and Russell, it was "ready to use" in logic-based systems,
- by stating that mathematics was only an example of language, it let think that it could be used for all the fields covered by human language. McCarthy was also interested in linguists that shared this idea of internal reference (Grice, Searle, etc.).

It is noteworthy that such an ontology is, as stated by Quine, a 'myth' that we are committed to believe or not, that we can disagree with, that we can compare with another one, but whose truth (or falseness) has little to do with experiments. This may explain why, in 1982, when Newell proposed to add a "Knowledge level" to logic-based

systems, Artificial Intelligence was still so far from real applications and tests (Rousseaux & Bouaziz, 2005). This could also explain why the introduction of ‘knowledge‘ in Artificial Intelligence brought both interesting application fields and theoretical confusion on what is an ontology.

Epilogue

This article was an attempt to contribute to the debate about the philosophical status of what is called ‘ontologies’ in the Semantic Web. We adopted a ‘genetic’ approach and had to go further a bibliographical dead-end in computer science to see if ‘ontologies’ could be connected to philosophical works, and in particular with Logical Positivism. Our result is that the missing link could be, very likely, Quine, a dissident of Logical Positivism, and McCarthy, the founder of Artificial Intelligence.

This result reveals that the term ‘ontology’ has then a very non-classic meaning: it has nothing to do neither with essence nor with experience. Such an ‘ontology’ confers to Artificial Intelligence (and then to the Semantic Web) a very speculative status, hardly compatible with real-world applications, except at the cost of dangerous theoretical trade-offs.

The questions that remain are why this filiation is not clearly assumed by explicit bibliographical references, why the Semantic Web promoters did not reveal what they owed to techniques, philosophy and people of the Artificial Intelligence domain. One could wonder whether it was not to avoid the arguments that are opposed to this domain for forty years (Dreyfus, 1972) and present this kind of approach as a dead end.

References

Aurélien Bénel, Elöd Egyed-Zsigmond, Yannick Prié, Sylvie Calabretto, Alain Mille, Andrea Iacovella, Jean-Marie Pinon (2001). Truth in the Digital Library: From Ontological to Hermeneutical Systems. *Proceedings of the fifth European Conference on Research and Advanced Technology for Digital Libraries* [ECDL], LNCS no. 2163. Berlin : Springer-Verlag. p. 366-377.

Aurélien Bénel, Sylvie Calabretto (2004). Ontologies... déontologie : réflexion sur le statut des modèles informatiques, *Linguistica Computazionale, n°20-21*. p. 31-47.

Tim Berners-Lee, James Hendler, Ora Lassila (2001). The Semantic Web. *Scientific American, May 2001*.

Jean-Pierre Cahier, Manuel Zacklad (2001). Expérimentation d’une approche coopérative et multipoint de vue de la construction et de l’exploitation de catalogues commerciaux « actifs ». *Document numérique n°3-4, Volume 5.* Hermes. p.45-64.

Jean Charlet, Bruno Bachimont, Jacques Bouaud, Pierre Zweigenbaum (1996). Ontologie et réutilisabilité : expérience et discussion. In: N. Aussenac-Gilles, P. Laublet & C. Reynaud (Ed.) *Acquisition et ingénierie des connaissances. Cepaduès Editions*.

Hubert Dreyfus (1972). *What Computers Can't Do.* New York: MIT Press.

Thomas R. Gruber (1991). The Role of Common Ontology in Achieving Sharable, Reusable Knowledge Bases. In: J. A. Allen, R. Fikes, & E. Sandewall (Eds.), *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*. Cambridge MA: Morgan Kaufmann. p. 601-602.

Thomas R. Gruber (1993). A translation approach to portable ontology specifications, In: *Knowledge Acquisition, Volume 5, Issue 2*. p. 199-220.

Nicola Guarino (1997). Some organizing principles for a unified top-level ontology, In: *Proceedings of AAAI 1997 Spring Symposium on Ontological Engineering*, AAAI Press.

Harry Halpin, Alexandre Monnin (2010). Tim Berners-Lee on PhiloWeb and Philosophical engineering, Interview. Available on Dailymotion.

Thomas S. Kuhn (1962). The structure of scientific revolutions, In: *International Encyclopedia of Unified Science, Volume 2, Issue 2.*

Jean-Louis Léonhardt (2008). *Le rationnalisme est-il rationnel ? L'Homme de science et sa raison.* Parangon.

Vladimir Lifschitz (1990). *Formalizing Common Sense: Papers by John Mc Carthy*. Norwood (New Jersey): Ablex Publishing.

Karl Popper (1934), *The Logic of Scientific Discovery (*translated from German in 1959). New York : Basic Books.

Willard Van Orman Quine (1948). On What There Is, *Review of Metaphysics.*

François Rastier (2010). Web Semantics vs the Semantic Web? The problem of keyness. In: M. Bondi, M. Scott (Eds.),*Keyness in Texts.* Amsterdam: John Benjamins Publishing. p. 93-110.

Francis Rousseaux, Thomas Bouaziz (2005). L'invention des Connaissances par les informaticiens. *Texto !, septembre 2005.*

Clay Shirky (2005). Ontology is Overrated: Categories, Links, and Tags. *O'Reilly Emerging Technology Conference [ETech'2005], San Diego, March 14-17, 2005*.

Monique Slodzian (2006). La terminologie : historique & orientations. Conférence invitée. *Semaine de la Connaissance, Nantes, 26-30 juin 2006.*

Barry Smith, Christopher Welty (2001). Ontology: Towards a new synthesis. *Proceedings of the International Conference on Formal Ontology in Information Systems*.

Barry Smith, Roberto Casati (1993). La physique naïve : un essai d'ontologie. *Intellectica*, 2, 17. p. 173-197.

John F. Sowa (1992). Semantic Networks. In: Stuart C. Shapiro (Ed.), *Encyclopedia of Articial Intelligence*, 2nd edition, vol. 2. New York: Wiley. p. 1493-1511.

John F. Sowa (2001). Signs, processes, and language games: Foundations for ontology. Invited talk. *Ninth International Conference on Conceptual Structures, "Broadening the Base", California, July 30-August 3, 2001.*

WEB METAPHYSICS BETWEEN LOGIC AND ONTOLOGY

Philippe Lacour
Institut d'Etudes Politiques de Paris; CIEPFC-Ecole Normale Supérieure de Paris
10 rue des Envierges
75020 Paris
+33 6 14 34 69 65
philacour@gmail.com

Abstract

One of the possible intersections between the Web and Philosophy lies in the use of the term 'ontology' by the Web architects. Indeed, the term "ontology" belongs to the classical vocabulary of the branch of philosophy called Metaphysics, which is concerned with the very nature of the world. Considering the Web as a form of (virtual) world, one could very well apply traditional philosophical questions to the stuff of this universe. Is it made of items (datas), processes (actions), or even things? What kind of ontology do we need to describe it? In this paper, we will argue that philosophy should focus less on ontology than on logic (namely, semantics) to tackle the issue, therefore slightly changing the way the problem is set. We shall take the case of Web Translation as an example. In so doing, we will show that a philosophy of the Web is justified to the extent that it somehow plays the role of psychoanalysis of culture, beyond the idea of a Critique (Kant) and of a psychoanalysis of knowledge (Bachelard).

Keywords

Ontology, Metaphysics, Semantics, Translation

INTRODUCTION

Whether the Web is a genuine philosophical object or just another fashionable way to recycle traditional concepts (the Web as a "rhizome"), authors ("Deleuze and the Web"), issues (is the Web a "real" world?), or disciplines (aesthetics, ethics, theory of knowledge) by applying them to a new (and shiny) domain can hardly be discussed *from a general point of view*. Rather, it might prove more efficient to tackle the problem through the investigation of a *precise* question (Web translation). In fact, such an inquiry will reveal that the metaphysics of the Web hesitate between two different possible orientations –let alone the normative aspects, in particular the legal ones (is the Web a public space? Do Web contents constitute a common good? [6]). Indeed, one might want to question the *ontological* nature of the Web, its "stuff" as a space of information and a virtual domain of action (digital environment): what is it made of? Items (datas, metadatas), processes (actions, operations)? What are their relations? But one might also question the relevance of this question, which leads to the current proliferation of ontologies. Isn't the focus on ontology the symptom of a misconceived theory of signification that links meaning to reference (datas) and inference (operations on datas, through metadatas)? A critical investigation into semiotics could show that the so-called ontological issue is in fact a semantic one, by unveiling an unconscious dimension of the metaphysical problem.

AN ONTOLOGICAL TURN?

The development of the Web has given way to many attempts of constitution of "ontologies", to the extent that their multiplication may give the impression of a permanent confusion, due to the uncertain nature of the Web individuals (documents, datas, resources?) [9]. Are ontologies the new battlefield of contemporary Web Metaphysics, or just the product of a rational misconception? Could a *critical* inquiry distinguish, in the concept of ontology, what concerns semantic networks (such as WordNet, for instance) and what concerns the traditional philosophical issue of the description of the world and of its majors elements (substances, categories, etc.)?

From a historical point of view, even if computer scientists declare that their 'ontologies' have nothing to do with the philosophical concept, we must note the analogy between their positions and the positions of the Logical Positivism in the 1930s [2]. Also, from a philosophical perspective, the link between semantics and ontology only seems obvious when word analysis implies the description of the world materials; that is: when signification is conceived as directly linked to reference (and sometimes inference, as opposed to difference), as it is the case in the logical and grammatical tradition [12]. For instance, for Aristotle, words refer to substances and accidents, according to a hierarchy that is to be discovered both in the language and the very stuff of being. Now, this lexical conception of meaning is precisely the one pervading the WordNet ontological project and its hierarchy between first, second and third order entities (from the Aristotelian list of categories, turned into a list of top concepts, to the individuals). The problem with this type of onto-logical classification –with the Porphyrian tree, and its wide use of the genus/species difference, as its paradigm– is that it rules out competing conceptions of meaning, upon which insists other linguistic tradition. Namely, the rhetorical, hermeneutical (and structuralist) approach underlines the importance of interpretation. If representation is the norm of language, then synonyms are a problem (how can two words have a different meaning, if they have the same reference?); but if interpretation is central (and not subsidiary) to the comprehension of meaning, then the context is to be taken into account, to the extent that there hardly exist any synonyms (they are individualized by their concepts). As a consequence, one must consider whether so-called "ontological" disputes concerning the Web mean actually anything more than semantic problems.

A LOGICAL (SEMANTIC) INQUIRY

As François Rastier argues [11], the Semantic Web, as originally designed by Tim Berners Lee after the model of formal ontologies, is a hierarchy of hierarchies. Its positivist vision of "datas" therefore only reproduces the above mentionned ambiguities of the referential conception of meaning. A proper semantics, on the contrary, could address the difficulties of information retrieval in another way. It would imply to conceive the semantic web as a social semantic (or a hermeneutical, or a pragmatic) one [13].

Interpretation, as a meaningful creation process, would not be conceived as secondary (in comparison with a set ontology) but as constitutive (along with a dynamic vision of signification) [12, 10]. Datas, which are supposedly given, neutral and non-interpretive, would be better understood as a complex construction. Instead of pretending to ground Web Semantics on a pile of standardized layers (as in the famous "layer cake"), one would rather imagine a flexible Semantics for the Web, consisting of a dynamic process including a document (with a testimonial value, submitted to description, revision and signature, according to a particular inquiry), interpretation (heuristic modelization), intersubjectivity (rational comparison of different points of view, organization of the conflict of interpretation) [1].

These remarks can be particularly well highlighted through the examination of the case of translation. And this is no coincidence. Indeed, one must insist on the importance of translation for the philosophy of language. It is an epistemological guide, which reveals the central role of interpretation in semantics, and which is all too often neglected by rigid positivist conceptions of meaning (e.g. in logical positivism, but also in the computationnalist view of cognition) [8].

WEB TRANSLATION AS A CASE STUDY FOR METAPHYSICS

According to Gilles-Gaston Granger, the "dream" of Contemporary Reason consists in the desire to create machines that could produce singularities [7]. For instance, machine translation would –allegedly– automatically produce individualized texts translated from an original source into various languages. Now, one has to distinguish between two cases of machine translation: the grammatical approach focuses on grammatical rules (as in the original Systran device); the statistical one on most frequent uses (a device made recently popular by Google Translate). However, both focuses on regularities rather than on singularities, which are equally important in linguistic productions, as the romantic and hermeneutic traditions clearly stated [4,5]. In order to get the best of the two worlds, many machine translation tools try to combine both regular approaches (as the new Systran). Still, they fail to reach the accuracy of human translation. Consequently, most computer assisted translation tools nowadays combine automatic translation with a human 'retouch' device, such as the Google Translation Toolkit, thus providing users with the ability to modify an automatically performed translation, while building their own translation memories [14]. From a critical perspective, such an evolution is actually not surprising, since it is obviously trying to make up for an all too often neglected dimension of language – its creativity.

As Rastier remarks, when people make fun of inadequate automatic translation, they fail to notice that this is less due to the performance of the machine than to the inappropriate underlying theory of translation [12]. Indeed, if the knowledge of rules is required in order to translate, it is however not sufficient. One does not translate from a language to another, but rather from a text to another one; and, in so doing, one does transform a system of norms into another one. And these norms cannot be reduced to the existence of grammatical or statistical rules, but they also include cultural categories, for instance: genres, styles, centuries, etc. For instance, in order to translate a play by Oscar Wilde into Arabic, one must construct an equivalent of the genre "theater" in the Arabic culture, where there existed no such thing before the colonial period.

One could therefore imagine to build a computer assisted translation device that would proceed in a different way, starting with human translation and using automatisation only for suggestions retrieval: such is the philosophy underlying the TraduXio project [16]. One does only use the machine to consult the concorder, browsing for relevant segments within a specific set of texts. Such a tool could help to build corpora in a reflexive and problem-oriented way, rather than in a merely quantitative perspective [4]. Not only does it illustrate the idea of a paradigm shift in Artificial Intelligence, switching from Machines that think to Machines that make people think [1]. But it actually stresses the crucial role of interpretation in semantics and, consequently, underlines that the issue at stake in "ontologies" is less a truly ontological than a *semantic* one.

CONCLUSION: WEB PHILOSOPHY AS PSYCHOANALYSIS OF DIGITAL CULTURE

The Web is a world of *meaning*, that is: of meaningful documents, and not a set of datas. It is therefore a complex universe that cannot be reduced to a composition of atomic items ("simple" datas), which could be organized in a rigid way (fixed "meta-datas") through the identification of robust standards. By unveiling the unconscious aspects of current ontological investigations, philosophy helps to convert Web Metaphysics from an ontological problem to a semantic one. The question switches from "What kind of world is a world of datas?" to "What is the meaning of these documents and corpora, according to what people do with them?" Hence the idea of a *socio-semantic* or *pragmatic* Web, based upon the idea of an interpretive semantics.

In so doing, philosophy does not criticize the power of reason in order to avoid confusion between representations and reality *per se* (phenomena and noumena); nor does it aim to limit the dogmatic use of reason by restricting its relevance to its empirical domain. Rather, it plays the role of a cultural psychoanalysis, beyond the traditional function of *critique* (in the Kantian sense) and the idea of a psychoanalysis of knowledge (Bachelard), which is tantamount to uncovering the unconscious that pervades the rational conception of the world and hinders objectivity. In the case of Web translation, the unconscious fantasy of contemporary reason (after Granger: to build machines that could produce singularities) clearly implies a misconception of translation and of semantics itself. Excavating this unconscious desire of our reason might help us to understand better our language, and therefore change our vision of the mission and scope of its automatic treatment.

REFERENCES

[1] Bachimont B., Artéfacture : des machines qui pensent aux machines qui donnent à penser, *Techniques & Culture* [En ligne], 23-24 | 1995, mis en ligne le 09 décembre 2005, Consulté le 29 février 2012. URL : http://tc.revues.org/571

[2] Bénel, A. 2012. Where do 'ontologies' come from? Seeking for the missing link" (forthcoming)

[3] Bénel, A., Lejeune, C. 2009. Humanities 2.0: documents, interpretation and intersubjectivity in the digital age, *Int. J. Web Based Communities*, Vol. 5, No. 4, pp. 562-576.

[4] Bénel, A. and Lacour, P. 2011. Towards a Collaborative Platform for Cultural Texts Translators. In Pierre Maret (éd.), *Virtual Community Building and the Information Society: Current and Future Directions.* Hershey (Pennsylvania): IGI Global.

[5] Lacour, P., Bénel A., Eyraud F., Freitas A., Zambon D. 2010. TIC, collaboration et traduction: vers de nouveaux laboratoires numériques de translocalisation culturelle, revue *Meta* 55(4), Presses de l'Université de Montréal.

[6] Lacour, P., Freitas, A., Bénel, A., Eyraud, F., Zambon, D., 2011. Translation and the New Digital Commons, In Proceedings of the Tralogy Symposium: "Translation Careers and Technologies: Convergence Points for the Future". DOI=http://lodel.irevues.inist.fr/tralogy/index.php?id=150

[7] Lacour, P. 2012. *La nostalgie de l'individuel. Essai sur le rationalisme pratique de Gilles-Gaston Granger*, Paris, Vrin.

[8] Lacour, P., Penser la traduction: entre Ricoeur et Granger, online Journal *Ricoeur Studies* (forthcoming)

[9] Monnin A. 2012, L'ingénierie philosophique entre rétablissement du pacte apophantique et dénaturalisation de l'ontologie (forthcoming).

[10] Rastier, F. 2012. *La mesure et le grain. Sémantique de corpus*, Paris, Honoré Champion.

[11] Rastier, F. 2004. Ontologie(s), in Revue des sciences et technologies de l'information, série: Revue d'Intelligence artificielle, vol. 18, n°1, p. 15-40.

[12] Rastier F. 2008, La traduction: interpretation et genèse du sens, *Texto !*, www.revue-texto.net/Lettre/Rastier_Traduction.pdf

[13] Zacklad M., Bénel A., Cahier J.P., Zaher L., Lejeune C., Zhou C. 2007. Hypertopic : une métasémiotique et un protocole pour le web socio-sémantique, in : Trichet F. (éd.) *Actes des 18ème journées francophones d'ingénierie des connaissances*, Cépaduès, Toulouse, pp. 217-228. http://publi.tech-cico.fr/files/zacklad_IC_07.pdf

[14] Google Translation Toolkit: http://translate.google.com/toolkit

[15] TraduXio: http://traduxio.hypertopic.org

CoReWeb: From linked documentary resources to linked computational resources

Alexandre Monnin
Université Paris 1 Panthéon Sorbonne
(Phico, EXeCO) /IRI/INRIA (Wimmics)
4 rue Aubry le Boucher
75004 Paris
Alexandre.Monnin@malix.univ-paris1.fr

Nicolas Delaforge
INRIA (Wimmics)
INRIA Sophia Antipolis Méditerranée
2004, route des Lucioles - BP 93
06902 Sophia Antipolis Cedex
Telephone number, incl. country code
Nicolas.Delaforge@inria.fr

Fabien Gandon
INRIA (Wimmics)
INRIA Sophia Antipolis Méditerranée
2004, route des Lucioles - BP 93
06902 Sophia Antipolis Cedex
Telephone number, incl. country code
Fabien.Gandon@inria.fr

ABSTRACT
The naive documentary model behind the Web (a single HTML Web page retrieved by a client from a server) soon appeared too narrow to encompass all cto account for dynamic pages, content negotiation, Web applications, etc. The Semantic Web raised another issue: how could we refer to *things outside* of the Web? Roy Fielding's REST style of architecture solved both problems by providing the Web its post-hoc "theory", making it a resource-oriented application. Recent evolutions (AJAX, HTML5, Linked Data, etc.) and envisioned evolutions (Web of devices, ubiquitous Web, etc.) require a new take on this style of architecture. At the core of the Web architecture and acting as a unifying concept beneath all its facets we find the notion of resource. The introduction of resources was very much needed for the Web to remain coherent; we now have to thoroughly redefine them to espouse its evolutions through time and usages. From the definition and the characterization of resources depends our abilities to efficiently leverage them: identify, publish, find, filter, combine, customize them, augment their affordance, etc.

Keywords
Web, resources, architecture of the Web, webarch, hypertext, hyperprocess, documentary resource, computational resource, rules.

1. INTRODUCTION
More and more often, the Web stands between us and the world. The Web of documents and data augments our perceptions of reality; the Web of applications and services, our grip on reality through the tasks we accomplish. It becomes at the same time both unavoidable in our daily activities and hardly manageable. On the Web, a resource is said to be anything and as the Web grows, everything around us is becoming a Web resource indeed.

This issue was already prevalent with the so-called Web of document. The first naive model behind the Web (a single HTML Web document retrieved by a client from a server) soon appeared too narrow to encompass all existing cases: dynamic pages, applets and scripts, content negotiation, Web applications, etc. The computational aspect, which at first appeared as an exception to the metaphor of the Web-as-a-universal-library, became the rule. In addition, the Semantic Web itself raised another issue: how could one refer to *things outside* of the Web? Roy Fielding's REST style of architecture solved both problems by providing the Web its post-hoc "theory", making it a resource-oriented application.

Recent evolutions (AJAX, HTML5, Linked Data, etc.) or even envisioned evolutions (Web of devices, ubiquitous Web, etc.) require a new take on this style of architecture. At the core of the Web architecture, acting as a unifying concept beneath all its various facets we find the notion of a resource. The introduction of resources was very much needed for the Web to remain coherent; we now have to thoroughly redefine them to espouse its evolutions through time and usages. From the definition and the characterization of resources depends our abilities to efficiently leverage them: identify, publish, find, filter, combine, customize them, augment their affordance, etc.

Justin Erenkrantz' definition of a resource as "a locus of computation" in his work on CREST (an computational update of REST) and the implications of plastering these loci all over the world around us will constitute our starting point in this article.

It also seems that among the different elements of the Web, the Web of linked data (i.e., linked meta-data or structured data) is to play an important role here. To manage the diversity of resources we can rely on another kind of diversity: the diversity of metadata. We believe that by overlaying a Web of semantic descriptions over the landscape of resources and by managing these linked data by the semantics of their linked schemas, the Web is giving itself a distributed and extensible paradigm to model its open pool of resources and to process these models. For this reason, we will lay the theoretical foundations of our work in this paper with the hope of getting closer to producing an ontology of resources based on Semantic Web formalisms in order to address many issues that are generally considered solely with regards to URIs.

2. WEB RESOURCES: TURNING THE PAGE OF THE DOCUMENTARY WEB

2.1 Giving names on the Web
In this first part, we wish to demonstrate that it is possible to account for the putative transition between a Web of document towards a Web of applications strictly *from an architectural point*

of view. Far from being just an historical account of the development of the interactive Web, with careful analysis of the introduction of JavaScript, the DOM, Ajax-based applications, etc., our endeavor will rather be one that aims to show that the basic concepts behind the Web and the constraints they entail were enough to undergo and even foster these evolutions.

At the heart of the original architecture of the Web [4] we find three basic concepts.

The first basic concept is the **URL** [6] or **URI** [7] [18]. Over time, the URI (Universal Resource Identifier) came to be thought of as a format of unique identifiers for naming and indicating any "resource" on the Web (this understanding of URIs stems from the REST style of architecture according to which parts of the Web were reinterpreted to cope with predicaments found in previous standards). If, in addition, such an identifier gives a path to obtain a representation of a resource, then it is also a URL (Universal Resource Locator) one of these famous Web "addresses" that everyone now knows about, even if, originally, they were not to be handled directly by users - e.g. "http://www.inria.fr/" is the URI(L) of INRIA home page. We could immediately note here that although these so-called addresses were initially not intended to be really used by humans, they are now part of our daily communications up to the point that there exists a market where they're valued and exchanged. Also, rather than addresses, which is actually a different concept, URLs should be understood as belonging to a subset of URIs, those URIs that are deferenceable. After all, URLs do not just locate representations, they retain the relation of identification between URIs and resources and add another relation, of access, to representations.

The second fundamental concept is the **HTTP protocol** [12] which allows for instance a client (e.g. a Web browser) to request a representation of the resource identified and "located" by a URL and get in return either the data of the resource representation or an error code indicating a problem, e.g. the famous 404 error indicating that the page you requested was not found. We should stress that the HTTP protocol does not only enable one to GET a representation but also to POST a new one, PUT an updated version or DELETE it.

The third fundamental concept was the **HTML** language to represent, store and communicate the representation(s) of the famous Web pages. It has ever since been complemented by other languages using an XML syntax to exchange any kind of structured data or document, one of the dialects of XML being a syntax for RDF, the linked data framework and core graph model of the Semantic Web.

All three basic concepts of the Web are especially important given that any current extension of the Web, including the **Web of data**, is fundamentally based on the first two concepts to identify the subject of data exchanged (URI) and transfer the data (HTTP). Indeed, the keystone of the architecture of the Web of data is the same as the classic Web: namely, the standard URI naming mechanism. However, unlike the documentary Web in which relationships are formed between anchors in hypertext documents, relationships in the Web of data are typed links (where types themselves are identified by URIs) between arbitrary resources (also identified by URIs). By relying on (HTTP) URIs for naming, on the HTTP protocol for data transfer, on the RDF graph model (instead of HTML) to describe and link resource, and on shared schemas, the recommendations of the Semantic Web outline an architecture for the world-wide interconnection of data sources and models.

2.2 Identity crisis

Yet, much work was needed to reach a shared agreement over the most basic building blocks of the Web. Standards for identifiers, for instance, evolved over time, from the first UDI draft [4] and URI specification [5] during the pre-W3C era (when the fundamentals of the Web were not yet clearly distinguished from their implementations) to the first standards concerning URLs and URNs (non-dereferenceable proper names), up until the latest URI RFCs. The work accomplished by Roy Fielding with the REST style of architecture [13] [14] was instrumental in reshaping the understanding both of Web identifiers and the HTTP protocol. It is also in Fielding's thesis that resources are defined for the first time. An immediate practical result consequence of REST was the fusion of what had previously been sundered between URLs and URNs back into URIs in 1997-1998 [29] [6]. French sociologist Laurent Thévenot [31] summarizes the agency of standards by explaining that they are "forms" that aim to generalize, extend, stabilize and equate a given technical reality. This is exactly what the Web achieved through REST and the recommendations it inspired.

Around the same time (19997-1998), other standards, the first explicitly dedicated to the Semantic Web, appeared. This conjunction is not really surprising considering that the Web had reached an unprecedented state of maturity. A "new" problem then seemed to arise. Formerly known as the httprange-14 [27] [19] – now issue-57 [20] – it consisted in understanding how one could distinguish between URIs that identify so-called "documents" and those that identify "things"[1]. This distinction itself was rephrased in terms of "information resources" (IR) and "non-information resources" (NIR) – with no real investigation with regards to whether or not these distinctions were tantamount to one another.

Basically, the httprange-14 may be summed up as an attempt to find the technical means to distinguish between IR and NIR by relying on the HTTP header sent by a server to a client in a typical HTTP negotiation. Actually, it is difficult to discuss the httprange-14 from a purely technical point of view since it has become marred with conflicting interpretation over time. What the httprange-14 actually says is that a 200 header will be followed by a representation, a 303 by a URI that identifies a second resource and is supposed to give access to a representation through a 200 header, and both 4XX and 5XX responses do not give access to anything.

Table 1: Summary of the Http-range14

HTTP code	Result	Indication
200 (OK)	(HTTP) representation	IR (and NIR?)
303 (See Other)	URI	Any kind of resource

[1] "Well, things and their descriptions are not the same and when people started using URIs to make assertions (using RDF, on the *Semantic* Web) they wanted to be able to say both http://cities.example.org/oaxaca has a radish festival every year on December 23rd and http://cities.example.org/metadata/oaxaca.html was written by Raphael Sabattini" [32].

4XX 5XX	Error message	Impossible to guess anything

One could infer, just by looking at the columns titles of **Table 1**, that the httprange-14 eschews in proving anything since the second column only contains HTTP-representations and URIs in the case of redirection (or error messages) while the third column, indicating what can be inferred from the previous one, is left open to interpretation. If resources are just "shadows" or "concepts" [14], then both information and non-information resource cannot be distinguished in terms of their potential accessibility: only representations being accessible by definition, not resources[2]. Hence, the first two rows of the third column will technically contain both IR and NIR.

However, this has not been the default interpretation. What the httprange-14 was supposed to provide was a clear separation between IR and NIR. The technical solution advocated failed to achieve that goal for the aforementioned reasons. Therefrom, a normative reading of the header responses was promoted instead of the more circumscribed technical solution first envisioned. Whenever a 200 header is served, says that reading, what we get is an IR. NIR are served indirectly, through a 303 header, by redirecting to an IR whose representations are then accessed by a client. The debate then focused on the relevance of this construal, mainly motivated by the need felt to determine whenever a URI identifies a document or a "thing" (our answer being that in both cases it identifies a *resource*, in accordance with the fundamentals of webarch).

Instead of just a technical relation, redirection thus became a good practice advocated in the publication NIR. While httprange-14 had completely failed as a purely technical tool, a normative reading was still possible. Many are still deterred by the difficulty of implementing redirection on a broad scale. That is why a new issue was opened by the TAG:

> At their meeting in 16th July 2007 [1] the TAG resolved to create a new issue, HttpRedirections-57, as a response to a community request [2] that we give further consideration to the use of the HTTP 303 status codes *and* other possible mechanisms of obtaining a description of a resource (typically a non-information resource) where the referenced resource is not capable of providing representations of itself.[3]

Echoing the AWWSW report, the issue really is about URI "definition", especially within RDF context:

> "When a URI appears in an RDF statement, how can the reader of that statement determine the author's intended meaning? What RDF triples characterize that meaning? Where does the meaning come from? How should the meaning be determined, particularly in the context of the HTTP protocol, for an http URI? Can we codify a suite of nose-following methods for semantic web use -- a recipe one can follow in order to obtain a canonical graph (or "definition", "description resource", "URI documentation") for a URI?"[4]

Rather than following that trail and search for additional ways of materializing the "meaning" of a URI, we would like to make sense of the existing Web by showing its fundamental coherence, accounting for both the Web of document and the Web of applications (current webarch discussions focusing more on the RDF side of things). This will require of a close examination of what is called a "resource", a task that can no longer be deferred for the purpose of reaching a solution.

3. ONLINE COMPUTATIONAL LOCI: FROM LIBRARIAN REFERENCES TO LOCUS OF COMPTATIONS

Everything was there from the start; in fact the Web was never purely documentary. At least if we are to take seriously the fundamentals of its architecture (and by doing so, lots of problem would simply not appear in the wild).

Looking at the definition of a resource, one can distinguish between three elements: a resource; the state of a resource; and the representational state. We shall examine each of these three elements in turns.

a) Resources

According to RFC 2396 [7], a resource can be anything. Roy Fielding called it a "shadow" or a "concept", thus making a strong distinction between resources and documents (even a digital one, understood, ultimately, as a binary set of 1 and 0 physically hosted somewhere). By definition, resources can never be accessed and are only manipulated through their representations (see [14], one section of paramount importance in their paper is fittingly entitled "Manipulating Shadows").

b) States of a resource

Resources have states. While resources remain the same (or at least *should*, since that is a normative statement which is contradicted on a daily basis), they also carry different results over time in terms of the representations that can be served to give information about them. One must thus distinguish between a resource and its state(s). This echoes the well-known distinction between *rules* and their *applications*. Alexandre Monnin [23] has previously suggested to understand resources as rules, thus specifying Fielding's claim that resources are concepts (it should be noted that concepts are often treated as rules in the philosophical literature). Assimilating the resource to a rule

[2] This is not always understood, as evidenced in RFC 3986 [18] where one can read "A Uniform Resource Identifier (URI) is a compact sequence of characters that identifies an abstract or *physical resource* [our emphasis]." No resource is physical *per se* yet if a resource is "*the semantics of what the author intends to identifies* [our emphasis]" [14], as defined in REST, then it can be said that what one intends to identify this way is a physical thing, though the resource itself won't. This distinction has been put forward to full scrutiny in order to better understand resources in [23] It is interesting to note that the httprange-14 was somehow theoretically fixed before the advent of the Semantic Web when as soon as the notion of resource was contrived to make sense of the Web departing from an understanding of "Web pages" as static documents. It is the focus on documents that again was the cause of the identity crisis. The fact that resources can be anything (as long as it is identified by a URI) also made it possible to build the Semantic Web on top of URIs as a mean to identify any resource (not just document) and link to it.

[3] Cf. [20].

[4] Cf. [3].

allows to better understand how and why states are produced. Basically, a resource generates states: over time (Web pages evolve, just as the result of search engine queries or application results in general) or punctually, through content negotiation (abbreviated as "conneg").

Of course, some cases seem at odd with this construal. Is Tim Berners-Lee a rule? Of course not. But a rule/resource being a means to identify Tim Berners-Lee, it will always depend on the way one individuates that "thing". It could be either "the founder of the Web", "the overall Director of the W3C" or "a man born of X and Y" (this is actually the Kripkean way of identifying people through across possible worlds despite the claim that rigid designators are adverse to definite descriptions), etc. Eventually, these are three different resources, or, in other words, three different objects, three different ways to pick-up something.

It is especially important make this distinction *since nothing warrants that a resource will adequately correspond to a "real thing" in the world simply because it has been published on the Web*; even more so since the goal of the Semantic Web is **not** to find a way out of this issue[5]. Resources need not always correspond to definite description but at least they must have enough content to specify what "an author intends to identify" [14]. This identification is thus possible by means of rules, corresponding to resources on the Web.

Even if the Semantic Web is to be conceptualized as a Web of "entities" (a characterization we borrow from the OKKAM project[6] [9], [30]), many of these entities are in fact the result of a complex publishing process that begins with people who edit Wikipedia and agree by consensus to identify something somehow. This is at least how DBpedia[7], one of the most successful applications of the Semantic Web, works.

We must accept once and for all the fundamentals of webarch. Fortunately, the architecture of the (Semantic) Web is no theory of truth. By contrast, it happens to be fuelled by a very different notion, *trust*. A paramount factor of trust is who the publisher of a resource is, whence the importance of *provenance* on the Web. All these elements, that were traditionally associated with the *epistemic* dimension of knowledge and dissociated from the *ontological* dimension, are now clearly intermeshed on the Web. For instance, as a telling fact that should not surprise us, it should be stressed that the definition of a resource given by Roy Fielding and Richard Taylor [14] doesn't shy away from mentioning the *intention* of an author – perhaps better described as one or more publishers in this context. A resource is thus always, at least partly, an intentional object, or rather what we'd call an *institutional object*, to better cope with the public nature of publication on the Web and its technical environment, both aspects corresponding to what is hereinafter referred to as the *editorial* and *computational* commitments.

c) The representational states of a resource

States remain abstract, just as resources, not accessible as such. What can be accessed is the HTTP-representation of the state of a resource. It can also be of various formats and many representations can be served for a given resource. While the latter need not be identical, they should at least be all faithful to a given resource. In other words, all of them must be *computable* as acceptable states (i.e., applications of the rule) of a resource (i.e., rule).

If my resource is "the original text of Shakespeare's MacBeth", a French translation in HTML will not do as faithful representation. This case illustrates a simple yet important fact: even the Web 1.0 was a Web of resources. Something that hasn't changed today, despite the advent of the Web of applications.

We thus adhere to Justin Erenkrantz' definition of resources as "loci of computation" as exposed in his work on the CREST style of architecture [11]. With a slight difference, since we also firmly believe that such a definition is true for the Web in general, not just the Web of applications. Erenkrantz' words fit very well within the general picture we try to draw where resources are rules when he uses the expression "network continuation" to describe them, thus underlying the dual aspect of stability and change[8] that essentially characterizes them.

4. WEB OF LINKED COMPUTATIONAL RESOURCE

It is commonly admitted to attach a version number to the Web, like 1.0, 2.0, 3.0, squared, etc., so that people may eventually come to think that there are several implementations of the Web. This is clearly misleading. Actually, we are still not using the full potential of what Tim Berners Lee had originally envisioned in the early nineties. In fact, rather than characterizing the Web,

[5] As Larry Masinter explains in a presentation entitled "Philosophy" of the Web", delivered at PhiloWeb 2012, WWW 2012 workshop in Lyon, France, http://www.slideshare.net/PhiloWeb/larry-masinter-philoweb): "Naming is printing money". One just has to remember that money can also be counterfeit, and the Semantic Web has not been designed to sort between genuine and counterfeit.

[6] http://www.okkam.org/

[7] http://dbpedia.org/About

[8] A major modification to this equilibrium would be the introduction of a new HTTP method to improve the red-write aspect of the Web, namely the PATCH one as described in a proposed standard, RFC 5789 [10]: "The PATCH method requests that a set of changes described in the request entity be applied to the resource identified by the Request-URI". Despite the lack of clear enough separation between resources and representations (for simplicity reasons and probably as an heritage of WebDAV conceptualization, though it certainly has the adverse consequence of partially excluding content negotiation. Rather than treat resources as modifiable files on a server, fitting mechanisms should be developed to apply updates on one kind of representations and spread it to others so as to preserve conneg). The idea behind this RFC and earlier proposals (including RFC 2068) is of great magnitude since it explicitly opens up the possibility that modifications applied to representations explain how resources may change over time (for instance, the URI identifying a given question about X on Stacked Overflow once it has been properly answered will thereafter identify a-question-about-X-that-has-been-answered-and-is-now-closed, prompting a very different attitude that translates in the set of potential actions made available to users by publishers (answering will no longer possible as perusing becomes the main available task, archiving will become the publisher's goal, etc.). In an interactive Web of applications prompting responses from users, such possibilities may become the norm, thus making it necessary to reassess what counts as "cool URIs".

these notations seem to betoken the (limited) grasp that we still have of it.

Up until recently, many industries were not ready to bring the Web to its full potential, nor were many computer scientists. Therefore when we talk about "documentary resources", one should really understand "the documentary application and understanding of Web resources".

Figure 1 emphasizes the differences between the first Web of documentary resources versus the current Web of computational resources. Of course, this distinction is mostly didactic. In practice, things are not so neatly and conveniently separated and, as we said, most of what we discover these days was here from the start. But for our purpose it is convenient to have a look at the "evolution of Web understanding". This approach allows us to highlight how much *practices* changed the structure of the Web.

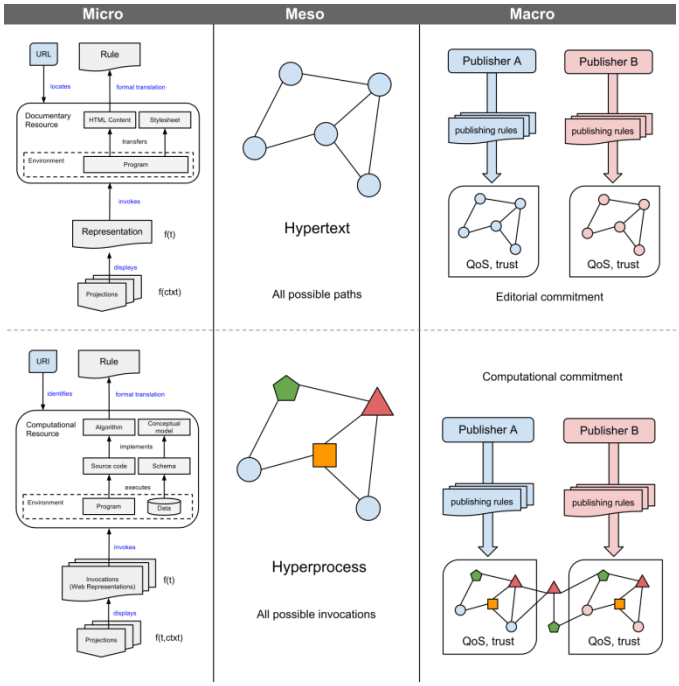

Figure 1: From Hypertext to Hyperprocess on a micro, meso and macro level.

In this section we propose a three-level analysis of the "Web of Computational Resource" (CoReWeb). The micro level focuses on the resource itself and its inner mechanisms. The meso level is about relations and interactions between computational resources. The macro level highlights the relations between the editorial policy of a publisher and the way he manages his Web resources.

4.1 Resources and other rules

Web resources are often published as part of bigger sets of resources that have in common to be named and managed by the same publisher. We consider that an *editorial policy* can be summarized as a structured rule set. Some of these rules are generic, others are specific and can inherit or be related to broader ones. From this, we assert that **any Web resource formally expresses the intersection of several of these publishing rules**. In other words, a Web resource is situated at the intersection of a number of publishing rules. A URI then gives access to a representational state that is the result of this intersection and its closure, while it is often perceived as identifying ony the most specific rule involved in generating the aforementioned representational state (otherwise known as "the" resource).

Indeed, the very way by which Web resources are cut out depends on their being distinguished from one another and included in a common set, an editorial ecosystem generally known as a "website" – even though such a notion bears little sense according to webarch. Actually, the set-theoretic approach, as found in the W3C recommendation POWDER ([1], [2]) allows to treat websites and RESTful Web services or data stores the same way: as "irisets" (in facts, sets of resources rather than IRIs, but the former are only manipulable as sets of IRIs[9] – groupings of resources identified by IRIs/URIs).

To borrow an analogy from linguistics, the "signified" in Saussure's theory is specified by relations of difference. By contrast, resources do share some common traits: they link to one another, to external resources, as mashups include parts of other resources, follow a given publishing policy being organized under specific categories, hierarchy, etc. Or, alternatively, in the case of Semantic Web resources, follow various axioms, share sets of properties and objects, etc. Yet, eventually, each must have a specific content distinguishing it from its neighbors. A resource is precisely this modicum atom of content that is supposed to remain stable, at least as much as possible, especially from a publisher's point of view, whereas representations as well as editorial policies do endure modifications (albeit allegedly much less often regarding the latter).

Here it may be useful to appeal to the distinction proposed by T.V. Raman [24], between "Web components" and "Web container":

> "(...) the need to provide a single point of access to oft-used information led to portal sites that aggregated all the information onto a single Web page. In this context, the various items of information can be viewed as lightweight Web components. The environment in which these components are hosted (such as the software that generates and manages the Web page) can be viewed as a Web container. Thus, common actions (such as signing in) were refactored to be shared among the various Web applications hosted by the Web container, a piece of software managing the user's browsing context."

Those rules reflect the editorial policy of a "website". For instance, this includes whether actions such as sharing content on a social network or using one's account to sign up or log in to a third-party website as well as being given the possibility to push the Facebook "like" button or Google's "+1" are made available. Such cases correspond to the integration of modular *components*, the grouping of which (and other editorial rules previously mentioned) gives rise to a Web *container*. Components and containers[10] may or may not be identified for themselves (a Facebook component might have one or more URIs while, by contrast, the decision to link a page to other pages "inside" a given container will not).

In any case, both containers and components are akin to *non-necessary rules* which add to a resource specific content enough

[9] See [9]: "A Resource Set is defined in terms of the IRIs of resources that are its members."

[10] We are using those words in a broader sense than as mere equivalents of "portlets" and "servlets". Many examples are given in section 4.1.2 ("Meso level").

details to compute concrete http-representations (the software used, HTML code, Web server headers and configurations, CSS style sheets, the JavaScript it includes, the JSP or PHP tags it uses, etc.). In other words, the policies or pieces of code that will generate a desired effect without belonging to the core-definition of a resource – i.e., without being confused with what a URI specifically identifies.

On the Web, attending to editorial policies and rules can either be done by one or many people. Since these tasks can be separated and often are in concrete situations, it is crucial to have them clearly distinguished from the inception.

4.1.1 Micro level

Technical evolutions have impacted both servers and clients. At the beginning, browsers were the only Web clients but now, we have many devices and applications that are able to connect to the Web and to get data and services from it.

Web servers were originally designed to propose a hypertext experience of "filesystem-like" remote services. Since the common gateway interface (CGI) their structure became increasingly complex. Nowadays, servers are able to negotiate with clients to adjust the response so that most of the content is generated on the fly. Any Web server is also compatible with at least one programing language that can trigger the processing of very sophisticated tasks that sometimes involve other remote services.

This point has important consequences on what is downloaded from those servers. One of the defined rationales behind documentary resources is that people have tried to preserve the *causal pathway between a reference and an informational content*, because it was constitutive of all our "real world" documentary reference systems. The "transition" from documentary resource to computational resource made more obvious that this artificially preserved causal relation had been broken. Now the downloaded content is what [17] called a "Web representation" of the resource, and can change each time a resource is invoked. The documentary location has been replaced by a *locus of computation*, or what we would call *a space of invocations***.**

Times at which "pages" were written with authoring tools like Adobe Dreamweaver or Microsoft Word now seem long gone. Today, blogs and wikis have permeated the Web and old-fashioned authoring practices are withering. From the server point of view, it is much more complicated to host a blog than a set of HTML files and CSS style sheets. Online editing tools involve scripting language capabilities, database and adequate security policies with possibly multiple ports opened to connect remote services, authentication API keys, etc.

To enable the Web of Data, the W3C have made slight changes in the specifications of Web architecture. URLs are now considered as (dereferenceable) URIs. From a linked data perspective, every URI minter/resource publisher is indeed strongly encouraged to make them dereferencable, so that it is possible to navigate between RDF concepts in the same manner as between pages. The 303 HTTP code is used to inform the "concept browser" that the resource he is asking for is not "informational". Hence, the technical distinction between Web pages, Web services and RDF concepts is no longer a valid one (our subsequent use of the received expression "Web pages" – or rather HTTP-representations – is entirely motivated by this observation).

URLs were initially locating documentary resources. CGI and REST have turn URLs into RPC passing parameters to scripts or web services. Now every URL is, and in a sense has always been, a URI. URI are identifying protean resources that can turn themselves in any format required by the client. Such are the *computational resources*.

Like with any program, to manipulate a computational resource, one has to *implement* an *algorithm* with a *programing language*, a *conceptual model* and *data*. Each of these parts has a strong impact on those Web representations a user can browse or a program parse.

As said before, a resource is a formal translation of necessary and non-necessary publishing rules but these rules themselves can change, the implementation can evolve to match a new technological context, a bug can be fixed, a new feature added, the database can also be updated with fresh data, etc. There are many reasons for Web representations to change and that is the real communicative power of the Web: an editor can instantly adapt the whole editorial chain synchronously in accordance with any informational or technological constraints.

The growth of Web communication in the last fifteen years resides mostly in the quickness with which information can pass from the state of data stored in one or more remote databases to a Web representation. Thus, the ease of update of the publication chain on a global scale induced by the architecture of the Web constitutes its greatest value and its biggest breach with previous editorial practices.

4.1.2 Meso level

As we have seen, through HTTP, any computational resource is likely to refer to other resources or communicate with them. This capability was exploited to add dynamicity and real-time content to Web pages, but it also has many applications in the Web of data.

4.1.2.1 Extending the pages communication capabilities

In 1995, Java applets were the first practical manner to asynchronously load remote content into Web pages. One year later, Microsoft introduced the *iframe* element designed to allow webmasters to include one Web page into another one. In 1999, the first XMLHTTP ActiveX control appeared with IE5. Now every browser proposes asynchronous communication capabilities and this technology, commonly known as AJAX for "Asynchronous JavaScript and XML", is very widely used.

Many widgets do use AJAX to connect a remote Web server and include real-time changes into the displayed content of a Web page. Real-time charts of stock exchange ratings, news tickers, Google maps, Google trends are just a few examples of applications using AJAX.

But with HTML5 and the brand new Websocket JavaScript API, things are going even further. Whereas AJAX is asynchronous (connections are closed after the server response is received), websockets provide persistent connection capabilities to Web pages, a feature that used to be characteristic of low level programming languages. Other evolutions like IndexedDB and WebGL APIs contribute even more to transform Web Pages into complex Web Applications [23]. Persistent connections enable the development of real-time applications, such as collaborative real-time painting or 3D games.

4.1.2.2 Public APIs, Dashboards, Widgets, Mashups

With the spread of Service-Oriented Architectures and the standardization of RPC (Remote Procedure Call) protocols, the

Web offers a wide pool of public services any Web developer can draw from to build innovative applications. These services can either be requested directly, or, more often, they provide widgets that should be integrated in Web pages.

Since 2005, many dashboard applications have emerged, like Netvibes[11], but quickly Google[12], Yahoo![13] and Microsoft[14] released their own dashboards providing a large variety of widgets like calendars, mail, contacts, todo lists, RSS readers, financial or weather survey tools.

Entire frameworks, like Life Ray[15] have been developed to build such platforms where the user can compose his own page made of portlets[16].

Now dashboard applications seem to wither in favor of more flexible widgets that can integrate into any page. It is impossible to reference them all here, so we will limit ourselves to some typical examples:

- *Data visualization*

Using either REST or SOAP protocols, it is now common to compose complex processing chains made of multiple remote service calls. The most typical combination is to provide a data stream to a visualization service and to integrate it into a Web page. As an example, one can mention Wordle[17], Many Eyes[18] or Google Maps[19].

- *Mashups*

A mashup is the result of the combination of several sources of information like RSS feeds. Yahoo! Pipes[20] is the best-known mashup application and his cousin, DERI pipes[21], includes semantic features. Other examples include 123People[22], a personal information aggregator and the Twitter API which gave birth to lots of applications like Bubble-T[23], Polemic Tweet[24]...

- *URL shortener*

With the Twitter's 140 characters restriction, URLs were often too long to be posted. To that purpose shortening services have appeared like TinyURL[25] and Bit.Ly[26]. Both provide a public API to get a short URL from a longer one. These very simple services are among the most used on the Web and within many Twitter clients.

- *Translation services*

For those who wish to get their Web page automatically translated in any language, Microsoft, Yahoo! and Google have published their solutions: Bing Translation API[27], Google Translation[28] and Yahoo! Babelfish translation service[29]. The final representation visualized by the user is thus the application of their web service to an initial Web representation that could itself call on many other Web resources.

- *Currency conversion*

It can be useful to delegate to a remote service the task of currency conversion according to current exchange rates. This is the purpose of web services like Exchange Rate API[30] or Open source exchange rates[31].

4.1.2.3 Web services orchestration and choreography

For people wishing to build much more complex services compositions from middleware architectures, to model Business processes as compositions of atomic tasks and to execute these compositions as single processes, several standards have been released by the W3C allowing what is called "service orchestration" and "service choreography"[32] [21] [28]:

> An orchestration specifies an executable process that involves message exchanges with other systems, such that the message exchange sequences are controlled by the orchestration designer. A choreography specifies a protocol for peer-to-peer interactions, defining, e.g., the legal sequences of messages exchanged with the purpose of guaranteeing interoperability. Such a protocol is not directly executable, as it allows many different realizations (processes that comply with it).[33]

Therefore resources are not only related to each other by navigation or composition links. They are nested into a much more complex interaction network mostly based on remote procedure calls and data exchange between servers. Consequently, qualifying the Web as a hypertext seems a little bit outdated. That is why we would prefer the term *hyperprocess* (actually, REST, by turning webarch into a resource-oriented architecture already

---

[11] http://www.netvibes.com/

[12] http://www.google.com/

[13] http://my.yahoo.com/

[14] http://live.com

[15] http://www.liferay.com/

[16] http://www.jcp.org/en/jsr/detail?id=286

[17] http://www.wordle.net/

[18] http://www-958.ibm.com/software/data/cognos/manyeyes/

[19] https://maps.google.com/

[20] http://pipes.yahoo.com/pipes/

[21] http://pipes.deri.org/

[22] http://www.123people.com/

[23] http://dev.fabelier.org/bubble-t/

[24] http://polemictweet.com/

[25] http://tinyurl.com/

[26] https://bitly.com/

[27] http://msdn.microsoft.com/en-us/library/ff512419.aspx

[28] https://developers.google.com/translate/v2/getting_started

[29] http://babelfish.yahoo.com/free_trans_service

[30] http://www.exchangerate-api.com/howto

[31] http://josscrowcroft.github.com/open-exchange-rates/

[32] In [15], we find an attempt to account for the client-server dialog mechanism both in the context of Web pages and Web services in a logical way in order to type the processes involved; in other words, so as to be able to determine whether "two processes that interact may be checked *before* the interaction". The Curry-Howard correspondence ensures that these logical types correspond to Web processes. While especially relevant to our own computational approach, by treating all URIs as URLs, and URLs as pointers in computing languages, it has the severe drawback of being oblivious to the fact that the Web is a publishing platform whom identifiers have two functions, none of which can be ignored.

[33] Source : http://en.wikipedia.org/wiki/Business_Process_Execution_Language#The_BPEL_language

had the immediate effect of discarding this notion of hypertext making it fully inappropriate for the Web; despite the enduring popularity of the word, it remains largely deprived of meaning in this context).

4.1.3 Macro level

On the one hand, and fortunately for Web users, the increasing complexity of server infrastructures, was progressively outsourced under the responsibility of specialized companies that provide hosting and administration services at low cost. The improvement of virtualization and monitoring technologies has also greatly simplified such system administration tasks.

On the other hand, it is more and more difficult for publishers to ensure a good *quality of service* throughout the entire processing chain. The technological stack and the processes involved in publishing a resource have become so complex and so distributed that it is becoming harder and harder to ensure a strict editorial commitment because as the Web grows in diversity, this commitment has turned into a computational one.

From the societal point of view, content publishers whose main activity was to produce content and to guarantee the quality of information now have to deal with various new constraints owing to the specificity of the medium. Beyond the increasing rate of publication, publishers must also face new stringent public expectations in terms of technical quality of service and interoperability.

Facebook, Twitter, Delicious and Google have imposed their "social ranking" tools ("I like" button, Google "+1", "Retweet") to publishers who must embrace these technologies otherwise they risk losing customers. Publishers must also consider the growing number of devices that people use to access information: smartphones, tablets, Kindle, television... The outsourcing of network infrastructures and servers adds another intermediate in the decision chain, which further complicates delivering a good quality of service. Browsers now even include calls to the cloud to delegate part of the rendering...

In summary, the gradual evolution from *hypertext* to *hyperprocess* has progressively added to the constraints of an *editorial commitment* those of a *computational commitment*.

5. CONCLUSION – TOWARD UBIQUITOUS HYPER-RESOURCES

The Web was already very rich with regards to the variety of the multimedia resources it hosted and linked to, and this richness is still increasing. With the advent of the mobile Web and the Internet of Things, we are going toward Web-augmented reality, ubiquitous Web and a Web of things or objects.

But while the Web is augmenting our reality, the objects and places of our lives, the latter are in turn increasing the number and variety of Web resources. This evolution will come with a price, namely an increase in the complexity of Web resources and their dependencies.

The architecture of the Web of data and the models of the Semantic Web may provide a way to match the diversity of online resources by means of a framework of metadata designed to annotate Web resources and exploit the semantics of their schemas to process them intelligently. Metadata and their schemas could be the keystone of the new resource-centric Web applications, their integration and interoperability.

It is conceivable that tomorrow, he who controls metadata on the Web controls Web resources, and through them a lot of things.

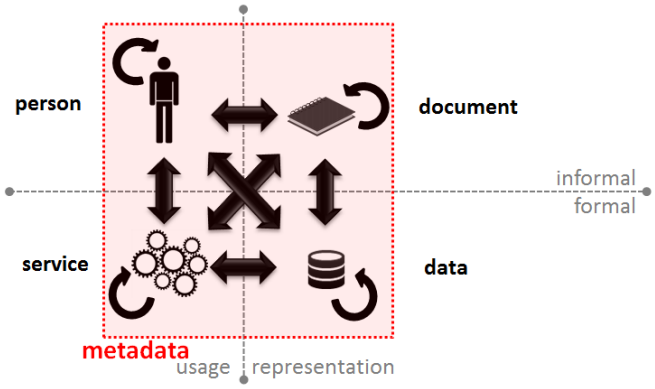

Figure 2: Synthetic view of the resource-centric Web architecture and the cross-cutting importance of metadata (as found in [16]).

6. REFERENCES

[1] Archer, P., Smith, K., and Perego, A. (Eds.). Protocol for Web Description Resources (POWDER): Description Resources. W3C (Sept. 1 2009). Retrieved from http://www.w3.org/TR/powder-dr/#operational

[2] Archer, P., Perego, A., and Smith, K. (Eds.). Protocol for Web Description Resources (POWDER): Grouping of Resources. W3C (Sept. 1 2009). Retrieved from http://www.w3.org/TR/powder-grouping/

[3] "Architecture of the World Wide Semantic Web" Task Group. *AwwswFinalReport*, W3C Wiki, (2012). Retrieved from http://www.w3.org/wiki/AwwswFinalReport

[4] Berners-Lee, T., Groff, J. F., and Cailliau, R. 1992. Universal Document Identifiers on the Network. *CERN, February*.

[5] Berners-Lee, T. 1994. RFC 1630 - Universal Resource Identifiers in WWW: A Unifying Syntax for the Expression of Names and Addresses of Objects on the Network as used in the World-Wide Web (June 1994). Retrieved from http://tools.ietf.org/html/rfc1630

[6] Berners-Lee, T., Masinter, L., and McCahill, M. 1994. RFC 1738 - Uniform Resource Locators (URL). (Dec. 1994). Retrieved from http://tools.ietf.org/html/rfc1738

[7] Berners-Lee, T., Fielding, R. T., Irvine, U. C., & Masinter, L. (1998, August). RFC 2396 - Uniform Resource Identifiers (URI): Generic Syntax. Retrieved from http://tools.ietf.org/html/rfc2396

[8] Berners-Lee, T., Cailliau, R., and Connolly, D. 2000. A Little History of the World Wide Web. *w3.org*. Retrieved from http://www.w3.org/History.html

[9] Bouquet, P., Stoermer, H., and Vignolo, M. 2011. Web of Data and Web of Entities: Identity and Reference in Interlinked Data in the Semantic Web. *Philosophy & Technology*. DOI:10.1007/s13347-010-0011-6

[10] Dusseault, L., & Snell, J. 2010. RFC 5789. PATCH Method for HTTP. IETF (March 2010). Retrieved from http://tools.ietf.org/html/rfc5789

[11] Erenkrantz, J. R. 2009. *Computational REST: A New Model for Decentralized, Internet-Scale Applications* (PhD Thesis). University of California Irvine. Retrieved from file:///F:/Articles/Erenkrantz-2009-CREST-Dissertation.pdf

[12] Fielding, R. T., Gettys, J., Mogul, J. C., Frystyk, H., Masinter, L., Leach, P., and Berners-Lee, T. (Eds.). 1999. RFC 2616 Hypertext Transfer Protocol -- HTTP/1.1 (June 1999). Retrieved from http://tools.ietf.org/rfc/rfc2616.txt

[13] Fielding, R. T. 2000. *Architectural Styles and the Design of Network-based Software Architectures* (PhD Thesis). University of California, Irvine. Retrieved from http://www.ics.uci.edu/%7Efielding/pubs/dissertation/fielding_dissertation.pdf

[14] Fielding, R. T., and Taylor, R. N. 2002. Principled design of the modern Web architecture. *ACM Transactions on Internet Technology (TOIT)*, *2*(2), 115–150.

[15] Fouqueré, C. 2011. Ludics and web: another reading of standard operations. In A. Lecomte and S. Tronçon (Eds.), *Ludics, Dialogue and Interaction*, Lecture Notes in Computer Science (Vol. 6505, pp. 58–77). Springer.

[16] Gandon, F., Faron-Zucker, C., and Corby, O. 2012. *Le web sémantique - Comment lier les données et les schémas sur le web ?* Dunod.

[17] Halpin, H., and Presutti, V. 2011. The identity of resources on the Web: An ontology for Web architecture. *Applied Ontology*, 6(3), 263–293.

[18] Hinden, R., Carpenter, B., and Masinter, L. 1999. RFC 3986 - Uniform Resource Identifier (URI): Generic Syntax. (1999, December). Retrieved from http://tools.ietf.org/html/rfc3986

[19] ISSUE-14: What is the range of the HTTP dereference function? - Technical Architecture Group Tracker. (n.d.). Retrieved from http://www.w3.org/2001/tag/group/track/issues/14

[20] ISSUE-57: Mechanisms for obtaining information about the meaning of a given URI - HttpRedirections-57 - Mechanisms for obtaining information about the meaning of a given URI. (n.d.). Technical Architecture Group Tracker. Retrieved from http://www.w3.org/2001/tag/group/track/issues/57

[21] Nickolas Kavantzas, David Burdett, Gregory Ritzinger, Tony Fletcher, Yves Lafon, and Charlton Barreto (Eds.). 2005. Web Services Choreography Description Language Version 1.0. W3C Candidate Recommendation 9 November 2005. W3C. Retrieved from http://www.w3.org/TR/ws-cdl-10/

[22] Monnin, A. 2012. L'ingénierie philosophique comme design ontologique : retour sur l'émergence de la « ressource ». *Réel-Virtuel*, *4*. Retrieved from http://reelvirtuel.univ-paris1.fr/index.php?/revue-en-ligne/3-monnin/

[23] Monnin, A. 2012. The artifactualization of reference and "substances" on the Web. Why (HTTP) URIs do not (always) refer nor resources hold by themselves. *American Philosophical Association Newsletter*, *11*.

[24] Raman, T. V. 2009. Toward 2 W, beyond web 2.0. *Communications of the ACM*, *52*(2), 52–59.

[25] Raman, T. V., & Malhotra, A. (Eds.). 2011. Identifying Application State - TAG Finding 01 December 2011. W3C. Retrieved from http://www.w3.org/2001/tag/doc/IdentifyingApplicationState

[26] Rees, J. 2012. The TAG Member's Guide to ISSUE-57 Discussion. *w3.org* (2012, March 27).. Retrieved, from http://www.w3.org/2001/tag/2012/04/57guide.html

[27] Rhys Lewis (Ed.). 2007. Dereferencing HTTP URIs (Draft Tag Finding 31 May 2007). W3C (May 31 2007). Retrieved from http://www.w3.org/2001/tag/doc/httpRange-14/2007-05-31/HttpRange-14

[28] Ross-Talbot, S., and Fletcher, T. (Eds.). 2006. Web Services Choreography Description Language: Primer. W3C Working Draft 19 June 2006. W3C (June 19 2006). Retrieved from http://www.w3.org/TR/2006/WD-ws-cdl-10-primer-20060619/

[29] Sollins, K., & Masinter, L. 1994. RFC 1737 - Functional Requirements for Uniform Resource Names (Dec. 1994). Retrieved from http://tools.ietf.org/html/rfc1737

[30] Stoermer, H. 2008. *OKKAM: Enabling Entity-centric Information Integration in the Semantic Web* (Thesis). DIT - University of Trento (Jan. 1 2008). Retrieved from http://eprints.biblio.unitn.it/archive/00001394/

[31] Thévenot, L. 1986. Les Investissements de Forme. In L. Thévenot (Ed.), *Conventions économiques*, Cahiers de Centre d'Etude de l'Emploi (pp. 21–71). Paris, PUF.

[32] Thompson, H. S. 2012. An introduction to naming and reference on the Web. HTML presented at the Philosophy of the Web seminar, La Sorbonne, Paris. (Jan. 28 2012). Retrieved from http://www.ltg.ed.ac.uk/~ht/PhilWeb_2012/

Containing the Semantic Explosion

Jaimie Murdock
Indiana University
Cognitive Science Program
810 Eigenmann Hall
Bloomington, IN 47405
jammurdo@indiana.edu

Cameron Buckner
University of Houston
Department of Philosophy
654 Agnes Arnold Hall
Houston, TX 77204
cjbuckner@uh.edu

Colin Allen
Indiana University
Cognitive Science Program
814 Eigenmann Hall
Bloomington, IN 47405
colallen@indiana.edu

ABSTRACT

The explosion of semantic data on the information web, and within digital philosophy, requires new techniques for organizing and linking these knowledge repositories. These must address concerns about consistency, completeness, maintenance, usability, and pragmatics, while reducing the cost of double experts trained both in ontology design and the target domain. Folksonomy approaches address concerns about usability and personnel at the expense of consistency, completeness, and maintenance. Upper-level formal ontologies address concerns about consistency and completeness, but require double experts for the initial construction and maintenance of the representation. At the Indiana Philosophy Ontology (InPhO) Project, we have developed a general methodology called *dynamic ontology*, which alleviates the need for double experts, while addressing concerns about consistency, completeness and change through machine learning over a domain corpus, and concerns about usability and pragmatics through human input and semantic web standards. This representation can then be used by other projects in digital philosophy, such as the Stanford Encyclopedia of Philosophy (SEP) and PhilPapers, along with resources outside of digital philosophy enabled by the LinkedHumanities project.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: User/Machine Systems ; I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods—*representations, semantic networks*

1. INTRODUCTION

"Information explosion" and "semantic web" are metaphors that have become clichés. Like many popular phrases they capture some important aspects of the situation, while disguising others. There has been rapid growth in the availability of both new and old materials on the Internet. The result of this rapid expansion is not, however, the pile of shrapnel that "explosion" might suggest. Rather, there is a highly linked set of pieces captured by the phrase "semantic web", which represents the connectivity but suggests a rigid approach to meaning that has fueled skepticism, and obscures the extent to which the possible semantic relations themselves grow exponentially as the number of linked sources increases. Thus, we prefer to characterize the situation as a "semantic explosion in the information web". This semantic explosion constitutes perhaps the most challenging problem that automatic methods for dealing with the information web must face.

Consider the problem of linking concepts as they occur in one philosophical resource to the concepts in another. Take, for instance, the term "realism". It is not adequate to treat the term as a proxy for a concept and link every occurrence of "realism" to every other because the term's meaning is context-sensitive according to whether it occurs in ethics, metaphysics, or political philosophy (among others), not to mention even finer contexts right down to the level of individual passages by authors who may use terms idiosyncratically. Even if one has disambiguated the term within a given digital resource or project, there remains the problem of how to link the disambiguated senses to occurrences of the term in other digital philosophy projects. As the web of information grows, so too do the interactions among its parts. So, noticing that realism is contrasted with idealism in some contexts (or databases), but with anti-realism in others, leads one to the question of whether the relationship between idealism and anti-realism is synonymy, and whether idealism in political philosophy is at all related to idealism in metaphysics. Furthermore, in trying to connect "realism" as that term is disambiguated in the Indiana Philosophy Ontology, to that term as it appears in other sources of philosophical content, such as the refereed journal articles covered in the PhilPapers database or the crowd-sourced Wikipedia entries on philosophical topics, there is a huge challenge in determining which "realism" belongs with which.

An appealing idea, at this point, is to regiment all of this into one overarching computational ontology that precisely fixes all the possible meanings. We think, however, that this one-size-fits-all approach is unlikely to succeed for a number of reasons. There may be pragmatically useful alternative ways of representing relationships among the data.[3] While we agree that a wide variety of digital humanities projects can benefit from the development of taxonomic schemes that make use of certain techniques of the computational ontologist, there are important differences between ontology design for the humanities and the approaches favored in other areas. Ontology science has evolved in large part to suit the needs of large projects in medicine, business, and the natural sciences. These domains share a cluster of features: their underlying structures have a relatively stable consensus, projects are amply funded, and a primary goal is often to render interoperable large bodies of data. In these projects, the best practices often require hiring so-called "double experts" – knowledge modelers highly trained in both ontology design and the subject domains – to pro-

duce a representation in the early stages of a project which is optimally comprehensive and technically precise.

There is a cluster of digital humanities applications, however, for which these practices are not ideal. These involve projects with principles of open-access and domains without the ample funding of the natural sciences. Additionally, ontologies for domains in which structural understanding is controversial or constantly evolving and projects which utilize computational ontologies to enhance search or navigation through asynchronously-updated digital resources must account for the dynamic nature of their resources – whether it is in the underlying corpus or in the judgments of the experts providing feedback on domain structure. On the positive side, these areas often offer more opportunities to collect feedback from users who are domain experts (but who lack expertise in ontology design).

Digital philosophy has many of the features described in the previous paragraph, and different projects have pursued different approaches to taxonomizing the subject matter of the discipline. The InPhO project starts with a top-level structure — a basic taxonomy provided by the editorial structure of the SEP — and a list of key terms assembled from various sources. We then data-mine the text of the encyclopedia to derive statistical hypotheses about term-relatedness. These hypotheses are presented to experts in a simple question format. Their answers drive an automated reasoning system to populate the seed taxonomy. This approach contrasts with PhilPapers, which employs a folksonomy-type approach to classifying the articles in its database. Another contrast is provided by DBpedia, which uses automated tools to extract lists of concepts and philosophers from the markup language used in the crowd-sourced Wikipedia articles. Our goal is to align these various representations based on semantic information embedded in them and the texts on which they are based. Given a taxonomy of concepts, for example, nearest neighbors and other related terms can help determine whether "realism" at a given location in one structure should be mapped to the same term appearing in another. The methods need to be as automatic as possible so that they can continue to be deployed even as the structures provided by these different projects change as a result of new input.

2. ONTOLOGY DESIGN

Effective ontology design for the humanities faces a number of sometimes conflicting desiderata. Formal approaches emphasize consistency, and in scientific contexts this can be achieved by axiomatizing the meanings of the terms represented. However, in the humanities, term meanings are among the most contested aspects of the disciplines. Another desideratum is completeness of coverage, which is hampered, however, by inadequate techniques for automatic term discovery, and vagueness about whether certain terms even belong to a given discipline. Usability is yet another desideratum, but usability by whom or by what? Formal ontologies stress machine readability and reasoning. However, given the semantic complexity and context sensitivity of terms in the humanities, usability by humans may be a more appropriate goal. The context-sensitivity also raises pragmatic issues relative to the various audiences. In a scientific gene ontology, for instance, it is clear that the end users are experts in the field for whom a large degree of consensus exists. In the humanities, however, there is much less consensus among experts, and disagreement is even encouraged. Representations of humanities disciplines, including philosophy, need to allow for the range of interpretations that different users will provide. Finally, scholarship in the humanities consists of suggesting novel interpretations of existing texts, new arguments and criticism, and novel concepts, necessitating not just the addition of new materials to existing databases, but continuously contributing to the semantic explosion as these new approaches interact with the existing structures. It is essential to automate as far as possible the maintenance of any digital representation of philosophy, lest the existing structures become quickly obsolete and abandoned.

In the following sections, we review the main approaches to ontology design, folksonomy and formal ontology. We conclude by outlining the principles of our favored approach, which we call "dynamic ontology", which attempts to leverage the strengths of each approach in semi-automatically generating structured representations of target domains.

2.1 Folksonomy

Social web (Web 2.0) and semantic web research were, for a time, conducted largely independently. Indeed, initial explorations of social computing were driven by skeptics of the grand unifying vision of the semantic web (e.g. [37]), who explicitly proposed "folksonomy" as an alternative method. This mutual antipathy may not be surprising, as the two approaches seem to offer competing visions for the future of the Internet. Social web researchers devise ways to harness the "wisdom of the crowds" to structure web data around information obtained from collaborative social interactions between large numbers of amateur users. Semantic web researchers, on the other hand, emphasized the need for a technically precise backbone of formal ontologies developed by small groups of experts highly-trained in the best practices of ontology design. Cultural differences have further fueled misconceptions and misunderstandings between these two research communities, often leading them to regard one another with mutual skepticism — though influential researchers have now recognized that the two approaches are not only not in conflict, but can even be complementary [9].

Both approaches have had some striking successes. Web 2.0 applications like Wikipedia, Facebook, Del.icio.us, and Flickr have reshaped the way average users interact with the Web. A key strength of such approaches lies in their ability to obtain large amounts of information from unskilled volunteers and to combine information obtained from many different kinds of sources creatively.

Since Thomas Vander Wal coined the term *folksonomy* in 2004 [37], there has been a surge of research on the effectiveness of folksonomy (see review in [36]). The use of the term itself is not precise, but a *folksonomy* is usually regarded as particular kind of knowledge base, one resulting from or induced upon the vocabulary derived from the collective tagging of shared resources by users in an online community. Folksonomy as a method comes with many advantages — the collection and organization of tags is virtually free, and the population of the knowledge base with resources with community relevance is guaranteed. Reviewers of taxonomic approaches have been encouraged by research on the "Wisdom of the Crowds", believing that the precision and recall of emergent tag behavior, once stabilized, will be superior to alternative methods.

From the beginning, critics recognized that folksonomy would face a variety of serious challenges. Mathes (2004) noted that tagging-based approaches inherently faces the problems of ambiguity, inconsistent orthography, and the unnoticed synonomy "inherent to an uncontrolled vocabulary". Many have worried about the idiosyncratic nature of tagging (characterized as the "long tail" phenomena, which describes the tendency of tag distributions to have a large number of rarely-used terms), though some research has shown that a stabilization of terminology can be reached in a community after a small amount of initial tagging behavior[7, 11]. Other studies, however, have shown that individual tagging behavior can evolve over time[4] as users become more familiar with the resources, raising challenges of intra-user lexical stability — though such behavior can gravitate towards "netlingo" tags that are not suitable for many taxonomic purposes (e.g. "fail", "toread", and "yum" — see [16]).

Many tagging systems have components designed to facilitate the stabilization of vocabulary — del.icio.us suggests commonly used tags, for example. Experts have also worried about the shallow depth of the taxonomic schemes induced on tags — Quintarelli (2005) noted the their lack of hierarchy, together with the concomitant difficulty in scaling the method up to organize larger knowledge bases. If tags are freely submitted by users, one must also worry about simply invalid tags; a study by Stvilia and Jorgensen shows that 37% of Flickr tags used in the Flickr Commons Project were misspelled or otherwise invalid, though this could be reduced to 15.3% with some simple pre-processing rules. A further worry of Kroski (2005) is that folksonomies are subject to "gaming"; because folksonomy systems often treat each user as an equal peer, they are vulnerable to "unethical users" who might "propagate tags ... in order to corrupt a system" (as a result, such systems would be wise to exploit user provenance data — e.g. see the ExpertRank system of John and Seligmann 2006 , and see Koutrika et al. 2007). Some have suggested that training users in tagging might help mitigate some of these problems [10], though other research has shown that users often balk at such training [19] and if financial incentives were required this approach would begin to incur the costs associated with double experts. In addition, Stvilia et al. 2011 found that the relationship between user tagging experience and perception of tagging quality is complex, with age and tagging experience being inversely related to the perceived suitability of tag-supplied terms, but Flickr familiarity and indexing experience having a positive relationship with term rating.

As clusters of tagging behaviors emerge, a further challenge is presented when one tries to use folksonomies to support tasks traditionally ascribed to ontologies — such as supporting reasoning and data interoperability. A variety of systems have been devised to leverage tagging libraries into ontologies, either using automated information-extraction or by designing tools which help users arrange tags in taxonomic relationships; but since tags are merely words applied to resources, these approaches face many of the same challenges that are faced by systems which attempt to learn taxonomies directly from text, including synonymy, polysemy, slang, inconsistent lexical forms and misspellings, and varying levels of generality. Marchetti et al (2007) have proposed that providing semantic support to taggers from resources like Wordnet and Wikipedia can help mitigate some of these challenges. Several approaches have been proposed to learn taxonomies and ontologies by using statistical techniques on tag distributions as a solution to this problem [28, 12, 1, 35], though all struggle with challenges posed by unregulated vocabulary and none offer the same rich level of structure as manually-encoded ontologies.

2.2 Upper-level Ontology

The grand vision of the formal "ontological" approach to the semantic web is to take a multi-layered approach to modeling reality. The task is divided into two levels: Lower-level domain ontologies are constructed to describe the entities of interest in specific domains; the types in the lower-level ontologies are then linked into a so-called "upper-level ontology", intended to describe the most basic, enduring features of reality. While by the nature of the method change is much more challenging on the formal ontological approach than with folksonomy, some of the largest formal domain ontology projects aspire to dynamism; the Gene Ontology project, for example, claims to offer "a controlled vocabulary that can be used for dynamic maintenance and interoperability between genome databases" [17]. Such dynamism is possible in the context of large biomedical informatics projects because they involve the efforts of very many dedicated biomedical informatics specialists working with manually designed taxonomies and ontologies. New data come pre-annotated because of the sophisticated equipment used for sequencing and other experiments. These features are only possible for deep-pocketed projects in domains studying relatively stable structure (though conceptual structure even in biology may not be so stable as one would think [8]).

The most significant challenges facing formal ontology are economic. Once elaborate and precise ontologies have been created, semantic web projects have faced the dilemma of either hiring expensive "double experts" to populate and maintain them or face inevitable data and user sparseness [3]. A further economic challenge is posed by the fact that projects developing domain-level ontologies are never sure which upper-level ontology should be linked to. Upper-level ontologies have now been an active area of research for fifteen years, and the diversity of choices appears to be increasing rather than decreasing. Modelers are now faced with a bewildering choice between a variety of inconsistent upper-level ontologies — including SUMO, DOLCE (and DnS), BFO, GFO, IDEAS, Cyc (and UMBEL), PROTON, OCHRE, and Sowa's [29, 20]. Debates in this area are bitter and protracted, given that there is often a significant commercial gains to be won by emerging as the "one ontology to unite them all". Many formal ontologists have by now abandoned the goal of selecting a single upper-level ontology [31], and recently attempts have been made to map the diverging upper-level ontologies into each other, such as COSMO (constructed largely out of categories from Cyc and SUMO) The most serious effect of these "ontology wars" has been that the population of elaborately-designed ontologies by the large amount of data already available on the web has languished while the battles are fought. Frustration with this process has in turn driven interest in alternative approaches to interoperability, such as the Linked Data initiative [2] which tries to obviate the need for upper-level ontologies by directly linking data in shared repositories (such as DBPedia and Freebase) .

More broadly, the debate over formal ontologies is situated

within a paradigm shift within artificial intelligence. The original vision of logic-based AI held that computers could display intelligence if only we could encode enough explicit expert knowledge into their systems. Though it quickly became apparent that this was a hopeless approach, the grand vision of "just getting enough knowledge" formally-specified continued to live on in the Cyc project for decades (as the largest remaining attempt in true artificial intelligence). Nowadays, even Cycorp has largely conceded this point — themselves turning away from grand visions of passing the Turing test with more specific practical goals, such as database translation. The push towards manual encoding and population of formal ontologies in the semantic web can further be seen as the last gasp of this knowledge-based approach to AI. Meanwhile, IBM's DeepQA system, showcased in Watson, starkly illustrates the lesson that, outside of a few specialized applications, it will simply never be practical to encode every scrap of knowledge in a clean, precise formal system [6]. Using a complex and heterogeneous system consisting of layers upon layers of diverse statistical methods, heuristics, partial ontologies, and ad-hoc tuning, the DeepQA methodology demonstrates that double experts are too expensive, and knowledge evolves too quickly for them to keep up with the problem at any hourly rate — especially in domains like the humanities. Practical intelligence, rather, requires a vastly more efficient tangle of statistical and ontological prowess, with both humans and computers contributing only what they do best.

3. DYNAMIC ONTOLOGY

At the InPhO project, we have developed a methodology for ontology population called *dynamic ontology*, which alleviates the need for double experts, while addressing concerns about consistency, completeness and change through machine learning over a domain corpus, and concerns about usability and pragmatics through human input and semantic web standards. Dynamic ontology follows a three-stage pipeline of data mining, feedback collection, and machine reasoning, summarized in Section 3.1. The core of dynamic ontology is the marriage of both human and computational resources in the design process. While human experts may be locally-blinded by their own familiarity with a subdomain, algorithmic processes can keep perspective over the entire corpus. Similarly, while data mining techniques may struggle with word sense disambiguation, human feedback can easily resolve such inconsistencies. This set of checks and balances helps maintain consistency in the resulting ontology.

As mentioned in the introduction, our pragmatic approach recognizes the likelihood that there is no single, correct view of the discipline. However, even if other projects do not agree with the InPhO's taxonomic projection, our statistical data and expert evaluations may still be useful for filling out alternative representations of the discipline. By exposing our data from each of the three steps in our procedure through an easy-to-use API, we enable the adoption of our system by other projects seeking alternative ways to construct meaningful and useful representations of the discipline. Additionally, by offering an open platform, we invite other projects to contribute relevant data and expert feedback to improve the quality of the service. By enabling linkages between the different representations, it becomes possible for end users to move among the different digital philosophy resources and make semantically interesting connections based upon their understanding of the concepts involved.

Thus, for example, the InPhO does not seek to replicate the cross-referencing structure of the SEP, but it provides data that the editors can use to select appropriate cross-references for the entries. Also, by providing links from each entry in the SEP to a dedicated InPhO page, readers can explore the concept network given by the InPhO representation, and use the InPhO portal as a way to discover other related resources outside the SEP, or to navigate back to related SEP articles via the InPhO taxonomy. Eventually, given fuller integration with PhilPapers, for example, it will be possible for end users and developers to navigate the conceptual space, the bibliographic network, and the linkages between specific thinkers using resources from all the various data providers.

3.1 The InPhO Workflow

Data Mining — Natural Language Processing (NLP) techniques are used over an external corpus (the SEP) to generate a lexicon of concepts and statistical hypotheses about semantic relevance and generality relations among various topics in the corpus. From this corpus we generate a co-occurrence graph in which each node represents a term in our set of keywords. An edge between two nodes indicates that the terms co-occur at least once. For each edge, the directed J-measure [32, 23] and conditional entropy [30] is calculated bidirectionally. The J-measure calculates the interestingness of inducing the rule "Whenever idea i is mentioned in a fragment of text, then idea j is mentioned as well" (for details see Niepert et al. 2007). This is used as a proxy for semantic relevance of term i to term j. While the J-measure can be used to estimate semantic distance, we are currently investigating alternative measures of semantic distance, as reviewed in Resnik (1999). We then apply to each node an informational metric of entropy. Entropy is used as a proxy for the generality of a each term, on the assumption that more general terms will have higher entropy. By combining these relevance and entropy measures, we obtain a directed estimate of hypernymy/hyponymy–the basic building blocks of taxonomies. Further details on data mining techniques can be found in [23].

Feedback Collection — The statistical hypotheses about hypernymy and hyponymy are presented to domain experts through online interfaces located both on the InPhO website and through the SEP editorial interface. Evaluations are presented as pairs of concepts, with a slider to indicate the relatedness of two terms, and a selection of whether the first term is more specific, more general, as general as, or incomparable with the second term. Users self-report levels of expertise when they sign up for our system, and each feedback fact is recorded with provenance information. This allows us to stratify feedback by self-reported education level and leverage expertise to resolve feedback inconsistencies. Experiments have been conducted on the effects of this stratification [22] and upon feedback collected from Amazon Mechanical Turk users[5]. Further details on feedback collection can be found in [25].

Machine Reasoning — User feedback is then combined with the statistical measures as the input for our machine reasoning program, which uses answer set programming to output a taxonomic view of the discipline. To reduce computational complexity, a *seed taxonomy* created by domain

experts is added to the input. Variations in the answer set program, the subset of user feedback used, the data mining techniques, or the seed taxonomy can allow us to generate different representations of the discipline. These variations can then be evaluated against the external corpus to find the most suitable population method [22]. Further details on the answer set programming techniques can be found in [24].

4. FUTURE DIRECTIONS

We are currently engaged in two projects that will contribute to the management of the semantic explosion on the information web that we described in the introduction. These projects will expand the range of digital philosophy projects and enable connections to other digital databases not solely concerned with philosophy. This expansion has at least two different fronts. On the one hand, as different pieces of the semantic and social webs become connected, the appropriate linkages between entities in these pieces need to be established. In the LinkedHumanities project, in collaboration with the University of Mannheim and jointly funded by the DFG in Germany and the NEH in the United States, we are exploring ways of matching entities across the various databases that contain semantic information about the concepts and major figures already represented by the InPhO. In the Digging by Debating project with partners in the UK, and joint funding from the NEH and JISC in the UK, we will be attempting to map the interactions among philosophy and the sciences across various timescales, using data from Hathi Trust, PhilPapers, and InPhO. The resulting tools will enable users to discover and represent arguments appearing both in historical texts and current articles.

5. ACKNOWLEDGMENTS

The work in this paper was funded by the National Endowment for the Humanities Office of Digital Humanities, Grant No. PW-50401-09.

6. REFERENCES

[1] D. Benz and A. Gotho. Position paper: Ontology learning from folksonomies. In *LWA'07: Lernen, Wissen, Adaption*, pages 109–112, 2007.

[2] C. Bizer and T. Berners-Lee. Linked data âĂŞ the story so far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22, 2009.

[3] C. Buckner, M. Niepert, and C. Allen. From encyclopedia to ontology: Toward a dynamic representation of the discipline of philosophy. *Synthese*, 182:205–233, 2011.

[4] T. Coates. Two cultures of fauxonomies collide, 2005. http://www.plasticbag.org/archives/2005/06/two

[5] K. Eckert, M. Niepert, C. Niemann, C. Buckner, C. Allen, and H. Stuckenschmidt. Crowdsourcing the Assembly of Concept Hierarchies. In *Proceedings of the 10th ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, Brisbane, Australia, 2010.

[6] D. Ferrucci, E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, and K. Aditya. Building watson: An overview of the deepqa project. *AI Magazine*, 31(3), 2010.

[7] S. A. Golder and B. A. Huberman. The structure of collaborative tagging systems. *Journal of Information Science*, 2005.

[8] P. Griffiths and K. Stotz. Experimental philosophy of science. *Philosophy Compass*, 3(3):507–521, 2008.

[9] T. Gruber. Ontology of folksonomy: A mash-up of apples and oranges. *International Journal on Semantic Web & Information Systems*, 3(2), 2005.

[10] M. Guy and E. Tonkin. Folksonomies tidying up tags? *D-Lib Magazine*, 12(1), 2006.

[11] H. Halpin, V. Robu, and H. Shepard. The dynamics and semantics of collaborative tagging. In K. Moller, A. de Waard, S. Cayzer, M. Koivunen, M. Sintek, and S. Handschuh, editors, *Proceedings of the 1st Semantic Authoring and Annotation Workshop*. CEUR-WS.org, 2006.

[12] H. Halpin, V. Robu, and H. Shepherd. The complex dynamics of collaborative tagging. In *WWW 2007*. ACM Press, 2007.

[13] A. John and D. Seligmann. Collaborative tagging and expertise in the enterprise. studying social tagging and folksonomy: A review and framework. *Journal of Digital Information*, 10(1), 2006.

[14] G. Koutrika, F. A. Effendi, Z. GyÃşngyi, P. Heymann, and H. Garcia-Molina. Combating spam in tagging systems. In C. Castillo, K. Chellapilla, and B. Davidson, editors, *AIRWeb '07: Proceedings of the 3rd International Workshop on Adversarial Information Retrieval on the Web*, pages 57–64. ACM Press, 2007.

[15] E. Kroski. The hive mind folksonomies and user-based tagging, 2005. http://infotangle.blogsome.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging/.

[16] C. Lu, J. Park, and X. Hu. User tags versus expert-assigned subject terms: A comparison of librarything tags and library of congress subject headings. *Journal of Information Studies*, 36(6):763–779, 2010.

[17] A. M, C. A. Ball, and J. A. Blake. Gene ontology: tool for the unification of biology. *Natural Genetics*, 25:25–29, 2000.

[18] A. Marchetti, M. Tesconi, F. Ronzano, M. Rosella, and S. Minutoli. Semkey a semantic collaborative tagging system. tagging and metadata for social information organization; a workshop at www2007, banff, alberta. In *Tagging and Metadata for Social Information Organization; a workshop at WWW2007*, 2007.

[19] K. Markey. Twenty-five years of end-user searching, part 2: Future research directions. *Journal of the American Society for Information Science and Technology*, 2(8):1123–1130, 2007.

[20] V. Mascardi, V. Cordi, and P. Rosso. A comparison of upper ontologies. Technical report, University of Genova Computer Science Department, 2006.

[21] A. Mathes. Folksonomies - cooperative classification and communication through shared metadata. Technical Report LIS590CMC, University of Illinois, 2004.

[22] J. Murdock, C. Buckner, and C. Allen. Evaluating Dynamic Ontologies. In A. Fred et al., editor, *IC3K*

2010, CCIS, pages 258–275. Springer-Verlag, Berlin Heidelberg, 2012.

[23] M. Niepert, C. Buckner, and C. Allen. A dynamic ontology for a dynamic reference work. In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, page 297. ACM, 2007.

[24] M. Niepert, C. Buckner, and C. Allen. Answer set programming on expert feedback to populate and extend dynamic ontologies. In *Proceedings of 21st FLAIRS*, 2008.

[25] M. Niepert, C. Buckner, and C. Allen. Working the crowd: Design principles and early lessons from the social-semantic web. In *Proceedings of Workshop on Web 3.0: Merging Semantic Web and Social Web 2009 (SW)^2*, Turin, Italy, 2009. CEUR Workshop Proceedings.

[26] E. Quintarelli. Folksonomies power to the people. ISKO Italy - UniMIB Meeting, 2005.

[27] P. Resnik. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of artificial intelligence research*, 11(4):95âĂŞ130, 1999.

[28] P. Schmitz. Inducing ontology from flickr tags. In *WWW2006: Collaborative Web Tagging Workshop*. ACM Press, 2006.

[29] S. K. Semy, M. K. Pulvermacher, and L. J. Obrst. Toward the use of an upper ontology for u.s. government and u.s. military domains: An evaluation. Technical Report MTR 04B0000063, The MITRE Corporation, 2004.

[30] C. E. Shannon. *A mathematical theory of communication*. University of Illinois Press, Urbana, Illinois, 1949.

[31] B. Smith. Ontology. In L. Floridi, editor, *Blackwell Guide to the Philosophy of Computing and Information*, pages 155–166. Blackwell, 2004.

[32] P. Smyth and R. Goodman. An information theoretic approach to rule induction from databases. *IEEE Transactions on Knowledge and Data Engineering*, 4(4):301–316, 1992.

[33] B. Stvilia and C. Jorgensen. Member activities and quality of tags in a collection of historical photographs in flickr. *Journal of the American Society for Information Science and Technology*, 61(12):2477–2489, 2010.

[34] B. Stvilia, C. JÃűrgensen, and S. Wu. Establishing the value of socially created metadata to image indexing. *Library & Information Science Research*, 2011.

[35] J. Tang, H. f. Leung, Q. Luo, D. Chen, and J. Gong. Towards ontology learning from folksonomies. In *IJCAI'09: Proceedings of the 21st international joint conference on Artificial intelligence*, pages 2089–2094. Morgan Kaufmann Publishers Inc., 2009.

[36] J. Trant. Studying social tagging and folksonomy: A review and framework. *Journal of Digital Information*, 10(1), 2009.

[37] T. V. Wal. Folksonomy definition and wikipedia, 2005.